

Luiz Eduardo Pizzinatto

**EXPER.TE: UM FRAMEWORK PARA
IDENTIFICAÇÃO TEMPORAL DA
EXPERTISE**

Florianópolis – SC

2019

Luiz Eduardo Pizzinatto

EXPER.TE: UM FRAMEWORK PARA IDENTIFICAÇÃO TEMPORAL DA EXPERTISE

Proposta de Trabalho de Conclusão de Curso
para obtenção do grau de Bacharel no curso
de Ciência da Computação na Universidade
Federal de Santa Catarina.

Universidade Federal de Santa Catarina – UFSC
Departamento de Informática e Estatística – INE
Graduação em Ciência da Computação

Orientador: Carina Friedrich Dorneles
Coorientador: Rodrigo Gonçalves

Florianópolis – SC

2019

Luiz Eduardo Pizzinatto

EXPER.TE: UM FRAMEWORK PARA IDENTIFICAÇÃO TEMPORAL DA EXPERTISE/Luiz Eduardo Pizzinatto. – Florianópolis – SC, 2019

183 p. : il. (algumas color.) ; 30 cm.

Orientador: Carina Friedrich Dorneles

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Santa Catarina - UFSC, 2019.

1. Identificação de expertise. 2. Identificação de especialistas 3. Recuperação da informação. 4. Expertise retrieval. 5. Expertise finding I. Carina Friedrich Dorneles. II. Universidade Federal de Santa Catarina - UFSC. IV. EXPER.TE: UM FRAMEWORK PARA IDENTIFICAÇÃO TEMPORAL DA EXPERTISE

Luiz Eduardo Pizzinatto

EXPER.TE: UM FRAMEWORK PARA IDENTIFICAÇÃO TEMPORAL DA EXPERTISE

Trabalho aprovado. Florianópolis – SC, 28 de junho de 2019:

Carina Friedrich Dorneles
Orientador

Rodrigo Gonçalves
Coorientador

Renato Fileto
Banca 1

Florianópolis – SC
2019

*“Theory and practice sometimes clash.
And when that happens, theory loses.
Every single time.
(Linus Torvalds)*

Resumo

Problemas de descoberta do conhecimento e busca da informação na web são desafios que ainda existem, mesmo com as diversas técnicas já desenvolvidas. Especificamente no contexto de Localização de Especialistas, as pessoas buscam por experts desde antes do uso do computador. Na web, um dos desafios dessa busca está em determinar automaticamente a expertise de uma pessoa a partir da análise dos diversos documentos disponíveis nas redes sociais, páginas pessoais ou em outros repositórios digitais, uma vez que essa informação dificilmente aparece escrita de forma direta. Como forma de facilitar esse processo, o presente trabalho faz uma proposta de framework capaz de determinar, com aprendizado não-supervisionado, os principais tópicos das produções de pesquisadores cadastrados na plataforma Lattes, levando em consideração tendências e mudanças que ocorreram ao longo do tempo. As palavras que compõem os tópicos encontrados são, posteriormente, usadas para inferir a especialidade de cada indivíduo. A análise temporal é adicionada com o objetivo de determinar a relevância dos trabalhos em relação ao período pesquisado e como a expertise do pesquisador evolui ao longo do tempo. O trabalho analisa alguns modelos pré-existentes que consideram o tempo para extração da expertise e, a partir de uma adaptação, implementa um framework com melhor aderência aos dados do problema em questão. Para avaliar a implementação e mostrar sua aplicação, um pequeno conjunto de currículos de pesquisadores é submetido ao framework, de maneira individual e, em um segundo momento, todos os pesquisadores pertencentes a um Programa de Pós-Graduação da UFSC são analisados em conjunto, onde tendências ao longo do tempo são discutidas e confrontadas com as linhas de pesquisa oferecidas.

Palavras-chave: identificação de expertise, identificação de especialistas, recuperação da informação, tempo, temporal expertise retrieval, temporal expertise finding.

Lista de ilustrações

| | |
|---|----|
| Figura 1 – Visualização de parte da página de um currículo cadastrado | 20 |
| Figura 2 – Resumo dos procedimentos usados durante o <i>expert finding</i> | 22 |
| Figura 3 – Estrutura do modelo de votação | 26 |
| Figura 4 – Representação visual do LDA | 29 |
| Figura 5 – Mudança dos tópicos de uma pessoa ao longo do tempo | 31 |
| Figura 6 – Análise temporal da expertise uma pessoa | 31 |
| Figura 7 – Exemplo de mudança de campos de pesquisa, com alteração em nodos mais altos | 33 |
| Figura 8 – Exemplo de troca de tópicos de interesse, dentro de um mesmo campo de pesquisa. | 33 |
| Figura 9 – Representação visual do TAT | 34 |
| Figura 10 – Camadas presentes no TET. | 36 |
| Figura 11 – Temporal random-walk | 38 |
| Figura 12 – Representação visual (no formato <i>plate notation</i>) do TOT, com o amos- trador de Gibbs. | 40 |
| Figura 13 – Representação do fluxo do framework | 42 |
| Figura 14 – Mapa de calor das palavras por tópico do currículo <i>a</i> | 62 |
| Figura 15 – Evolução dos tópicos ao longo do tempo do currículo <i>a</i> | 63 |
| Figura 16 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo <i>a</i> | 63 |
| Figura 17 – Mapa de calor das palavras por tópico do currículo <i>b</i> | 67 |
| Figura 18 – Evolução dos tópicos ao longo do tempo do currículo <i>b</i> | 67 |
| Figura 19 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo <i>b</i> | 68 |
| Figura 20 – Mapa de calor das palavras por tópico do currículo <i>c</i> | 72 |
| Figura 21 – Evolução dos tópicos ao longo do tempo do currículo <i>c</i> | 72 |
| Figura 22 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo <i>c</i> | 73 |
| Figura 23 – Mapa de calor das palavras por tópico de todos os currículos do <i>PPGCC</i> | 77 |
| Figura 24 – Evolução dos tópicos ao longo do tempo de todos os currículos do <i>PPGCC</i> | 78 |
| Figura 25 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, de todos os currículos do <i>PPGCC</i> | 79 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Distribuição dos currículos por área do conhecimento, conforme dados de julho de 2017 | 18 |
| Tabela 2 – Distribuição dos currículos por nacionalidade, conforme dados de julho de 2017 | 18 |
| Tabela 3 – Distribuição dos currículos por nível de formação, conforme dados de julho de 2017 | 19 |
| Tabela 4 – Notação utilizada pelo TOT | 40 |
| Tabela 5 – IDs dos currículos utilizados nos testes | 55 |
| Tabela 6 – Parâmetros utilizados na inicialização do TOT. | 56 |
| Tabela 7 – Parâmetros utilizados na inicialização do TOT na execução para os currículos do PPGCC. | 57 |
| Tabela 8 – Idiomas detectados no currículo <i>a</i> , com amostras de traduções realizadas | 60 |
| Tabela 9 – Tabela com estatísticas sobre os bigramas do currículo <i>a</i> | 61 |
| Tabela 10 – Exemplos de bigramas substituídos nos documentos do currículo <i>a</i> . . . | 61 |
| Tabela 11 – Exemplos de produções extraídas diretamente da página do currículo <i>a</i> | 62 |
| Tabela 12 – Idiomas detectados no currículo <i>b</i> , com amostras de traduções realizadas | 64 |
| Tabela 13 – Tabela com estatísticas sobre os bigramas do currículo <i>b</i> | 65 |
| Tabela 14 – Exemplos de bigramas substituídos nos documentos do currículo <i>b</i> . . . | 65 |
| Tabela 15 – Exemplos de produções extraídas diretamente da página do currículo <i>b</i> | 66 |
| Tabela 16 – Idiomas detectados no currículo <i>C</i> , com amostras de traduções realizadas (<i>parte 1</i>) | 69 |
| Tabela 17 – Idiomas detectados no currículo <i>c</i> , com amostras de traduções realizadas (<i>parte 2</i>) | 70 |
| Tabela 18 – Tabela com estatísticas sobre os bigramas do currículo <i>c</i> | 70 |
| Tabela 19 – Exemplos de bigramas substituídos nos documentos do currículo <i>c</i> . . . | 71 |
| Tabela 20 – Exemplos de produções extraídas diretamente da página do currículo <i>c</i> | 71 |
| Tabela 21 – Linhas de pesquisa do PPGCC e suas subdivisões. | 74 |

Lista de códigos

| | | |
|---|---|----|
| 1 | Exemplo de elemento que representa um artigo dentro do XML | 44 |
| 2 | Outro elemento que representa um artigo, extraído do mesmo XML. | 45 |
| 3 | Documentos gerados a partir dos artigos apresentados nos Códigos 1 e 2, respectivamente. | 45 |
| 4 | Definição de um item responsável por extrair os <i>artigos publicados</i> do XML de um currículo | 47 |

Lista de abreviaturas e siglas

| | |
|-------|---|
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| DTD | Document Type Definition |
| ER | Expertise retrieval |
| IR | Information retrieval |
| LDA | Latent Dirichlet Allocation |
| TAT | Temporal-Author-Topic |
| TET | Temporal-Expert-Topic |
| TOT | Topics over Time |
| XPath | XML Path Language |

Sumário

| | | |
|------------|---|-----------|
| 1 | INTRODUÇÃO | 13 |
| 1.1 | Motivação | 14 |
| 1.2 | Objetivo Geral | 16 |
| 1.3 | Objetivos Específicos | 16 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 17 |
| 2.1 | Plataforma Lattes | 17 |
| 2.1.1 | Estrutura de um currículo | 19 |
| 2.2 | Expertise Retrieval | 22 |
| 2.2.1 | Modelagem por tópicos | 23 |
| 2.2.2 | Modelos com grafos de relacionamento | 24 |
| 2.2.3 | Modelo de votação | 25 |
| 2.2.4 | Outros modelos | 27 |
| 2.3 | Tipos de aprendizado | 27 |
| 2.4 | Recuperação da expertise com o Latent Dirichlet Allocation: modelagem por tópicos com aprendizado não supervisionado | 27 |
| 2.5 | Temporalidade na identificação da especialidade | 30 |
| 3 | TRABALHOS RELACIONADOS | 32 |
| 3.1 | Modelo Hierárquico | 32 |
| 3.2 | Modelo Temporal-Author-Topic | 33 |
| 3.3 | Modelo Temporal-Expert-Topic | 34 |
| 3.4 | Modelo Random Walk | 36 |
| 3.5 | Modelo Topics Over Time | 38 |
| 4 | O FRAMEWORK EXPER.TE | 41 |
| 4.1 | Visão Geral | 41 |
| 4.2 | Concepção do framework | 42 |
| 4.2.1 | Fonte de dados, processamento de arquivos e geração dos documentos | 42 |
| 4.2.2 | Tradução | 49 |
| 4.2.3 | Limpeza e normalização | 49 |
| 4.2.3.1 | Limpeza Básica (Passo 4.a) | 50 |
| 4.2.3.2 | Normalização de palavras (Passo 4.b) | 50 |
| 4.2.3.3 | Tratamento dos bigramas (Passo 4.c) | 50 |
| 4.2.3.4 | Limpeza das palavras desnecessárias globais (Passo 4.d) | 51 |
| 4.2.3.5 | Limpeza individual por currículo (Passo 4.e) | 51 |

| | | |
|------------|--|------------|
| 4.2.3.6 | Limpeza após agrupamento (Passo 4.f) | 51 |
| 4.2.3.7 | Outras considerações | 51 |
| 4.3 | Implementação | 52 |
| 5 | EXPERIMENTOS E RESULTADOS | 54 |
| 5.1 | Objetivos | 54 |
| 5.2 | Metodologia | 54 |
| 5.2.1 | Experimentos com currículos individuais | 55 |
| 5.2.2 | Experimento com grupo de pesquisadores | 56 |
| 5.3 | Métricas de avaliação | 57 |
| 5.3.1 | Mapa de calor | 57 |
| 5.3.2 | Evolução dos tópicos | 58 |
| 5.3.3 | Probabilidade posterior acumulada | 58 |
| 5.4 | Análise dos resultados | 59 |
| 5.4.1 | Experimentos com currículos individuais | 59 |
| 5.4.1.1 | Currículo <i>a</i> , da área de Ciência da Computação | 59 |
| 5.4.1.2 | Currículo <i>b</i> , da área de Economia | 63 |
| 5.4.1.3 | Currículo <i>c</i> , da área de Ciência da Computação | 68 |
| 5.4.2 | Experimento para grupo de pesquisadores | 73 |
| 6 | CONSIDERAÇÕES FINAIS | 80 |
| 6.1 | Limitações | 80 |
| 6.2 | Trabalhos Futuros | 81 |
| 6.3 | Conclusão | 82 |
| | REFERÊNCIAS | 84 |
| | APÊNDICES | 87 |
| | APÊNDICE A – ESTRUTURA DO CURRÍCULO LATTES | 88 |
| | APÊNDICE B – TRECHO DE UM CURRÍCULO LATTES EM XML | 97 |
| | APÊNDICE C – STOPWORDS DA LÍNGUA PORTUGUESA | 99 |
| | APÊNDICE D – ARTIGO | 103 |
| | APÊNDICE E – FONTES | 117 |

| | |
|---|------------|
| ANEXOS | 142 |
| ANEXO A – ESTRUTURA DO ARQUIVO XML DE CADA CUR- RÍCULO | 143 |

1 Introdução

A busca pela informação está fortemente presente no cotidiano das pessoas, desde suas ações mais básicas. A simples necessidade de buscar um número de telefone, encontrar uma receita ou mesmo uma foto, seja em meio físico ou digital, requer o acesso a uma base de dados para que o item desejado seja recuperado. Nas últimas décadas, dada a crescente massa de informação disponibilizada pelas Tecnologias de Informação e Comunicação, a Ciência da Computação e Ciência da Informação vêm abordando e discutindo esse assunto para encontrar meios de facilitar o acesso aos dados (ALVES et al., 2007).

Manejada inicialmente apenas por especialistas, a informação passou a ser buscada e recuperada por diferentes usuários finais – primeiro em sistemas informatizados de recuperação da informação; mais tarde, via internet e *world wide web* (GIORDANO; BIOLCHINI, 2012). A *web* passou a ser considerada o maior acervo de informações do mundo e encontra-se em crescimento exponencial. Lidar com uma massa tão grande de dados a fim de encontrar algo relevante pode ser complicado, requerendo técnicas específicas para se obter sucesso.

A recuperação da informação, ou *information retrieval* (IR) foi definida formalmente por Mooers (1951) como o nome do processo ou método pelo qual um usuário potencial da informação é capaz de converter sua necessidade em uma lista real de referências para documentos armazenados contendo as informações úteis para ele. É o processo de busca ou descoberta em relação à informação armazenada. De uma maneira mais prática, na definição de Mata-Rivera et al. (2015), IR é um campo de pesquisa amplo e interdisciplinar, que inclui indexação da informação, classificação da relevância, motores de busca, medidas de relevância (como *precision e recall*) e seu principal objetivo é corresponder à necessidade de informação do usuário. Na maioria dos casos, a IR contemporânea está vinculada a algum sistema de informação, para facilitar as buscas através de uma interface amigável.

Dentro das necessidades de busca, existe uma específica que visa encontrar especialistas de uma determinada área ou assunto. Pessoas têm procurado por especialistas desde antes do advento dos computadores e estes podem ser demandados não somente para responder a questionamentos de sua área, mas também para serem designados em alguma função ou trabalho, inclusive dentro de organizações (BALOG; FANG; RIJKE, 2012). Isso é fácil de ser exemplificado na área médica: os médicos são profissionais especialistas na grande área da saúde, mas é comum que possuam uma sub-especialidade, como pediatria, neurologia, dermatologia, cardiologia, sendo o mais adequado escolhido conforme o tipo de problema a ser resolvido. Da mesma maneira, na área da tecnologia, existem especialistas nas áreas de segurança da informação, infraestrutura de redes, bancos de dados,

microcontroladores, entre várias outras. Encontrar o expert na área desejada é um dos grandes desafios da IR.

A busca por especialista é chamada na literatura de *expertise retrieval* (ER) (ou ainda *expertise location*, *expertise identification*) e se refere à área genérica de ligar pessoas a áreas do conhecimento e vice-versa (BALOG; FANG; RIJKE, 2012). Dentro da ER, as duas principais tarefas investigadas são: *expert finding* ("Quem são os especialistas no tópico X?") e *expert profiling* ("Quais são os tópicos que a pessoa Y conhece?"), sendo que a primeira recebeu consideravelmente mais atenção que a segunda (RYBAK; BALOG; NØRVÅG, 2014).

Devido à quantidade de informações disponibilizadas na web (de maneira semi-estruturada ou não), várias pesquisas foram desenvolvidas na última década em torno da ER. Desde então, o processo de encontrar especialistas vem sendo estudado em contextos variados, tanto no meio acadêmico quanto no industrial (RUSSELL-ROSE; CHAMBERLAIN, 2016), mas apesar dos avanços feitos até o momento, a questão de como prover acesso efetivo à expertise está longe de ser resolvida e continua a ser encarada de diferentes pontos de vista (BALOG; FANG; RIJKE, 2012). No passado, os frameworks usados na busca por especialistas poderiam ser majoritariamente divididos em duas categorias: aqueles baseados em grafos de conectividade (ligações entre documentos levando em conta citações, coautorias e casamentos de palavras-chave) e aqueles que fazem uso de uma camada *invisível* de tópicos latentes para capturar a relação semântica entre os documentos (DAUD et al., 2010). No caso dos grafos de relacionamento, geralmente a informação semântica interna dos documentos é ignorada (o conteúdo não é analisado) e, no caso das abordagens por tópicos, a semântica interna é capturada, mas ignora outras influências, como importância das conferências em que os documentos foram publicados e relações entre coautores. Além disso, a maioria das abordagens ignora a modelagem do tempo. Uma solução ideal na busca por especialistas deveria considerar todas as características que mudam rapidamente. Infelizmente, a maioria das abordagens existentes ignoram (ao menos parcialmente) as mudanças e tendências de diferentes informações (LI; TANG, 2008).

1.1 Motivação

Considerando que a *web* é extremamente dinâmica, ignorar o fator *tempo* não é uma suposição realista a ser feita, pois a maioria dos conjuntos de dados (como artigos de pesquisadores, blogs e notícias) não possuem co-ocorrência estática; muito pelo contrário, padrões nos dados coletados estão sempre mudando, mostrando tendências de crescimento ou diminuição ao longo do tempo (DAUD et al., 2009). Dizer que uma pessoa é especialista em um determinado tópico em 2018, com base em evidências de 2010, pode levar a

conclusões equivocadas.

Os meios acadêmico e corporativo quase sempre estão interessados em pessoas que possam ser consideradas especialistas em um determinado assunto no *momento atual*. No caso do meio acadêmico, pode ser interessante encontrar um professor ou pesquisador para:

- compor bancas de defesa de mestrado ou doutorado;
- criar questões ou compor bancas de concursos públicos;
- fazer parte de grupos de revisores de trabalhos submetidos em periódicos ou eventos científicos;
- compor uma nova linha de pesquisa dentro de um programa de pós-graduação.

Em seu trabalho, Daud et al. (2010) cita dois exemplos onde considerar a relevância do tempo é importante para encontrar pesquisadores: (i) um pesquisador pode ter seguido uma linha de pesquisa ao longo de sua carreira, publicando vários artigos sobre o tópico *microprocessadores*, mas, a partir de um determinado ponto no tempo, sua linha de pesquisa muda gradativamente e os tópicos das publicações não são mais os mesmos. No passado, este pesquisador poderia ser considerado um especialista em *microprocessadores*, o que pode não ser verdade no presente. Ainda assim, mesmo este especialista estando em hiato nas produções sobre o tópico especificado (ii) novos autores podem realizar coautorias com ele em tópicos semelhantes (ou iguais) a *microprocessadores*, retomando seu destaque neste assunto no período específico das coautorias.

No meio corporativo, a consideração da temporalidade é análoga. Mesmo que uma pessoa tenha trabalhado por muitos anos em uma determinada área, mudanças ocorridas em sua vida profissional podem acarretar em alterações nas suas especialidades. Por exemplo, um analista de sistemas que trabalhou por 10 anos com uma tecnologia e passou a atuar como gerente de projetos por outros 10 anos provavelmente não pode ser mais considerado um especialista na tecnologia em questão.

Especialistas no meio corporativo podem ser demandados para:

- compor equipes multidisciplinares;
- preencher uma vaga de trabalho;
- emitir pareceres;
- liderar projetos específicos;
- participar como palestrante em eventos;
- ministrar cursos de formação de equipes.

Fica claro, portanto, que o fator *tempo* é importante na determinação da relevância dos trabalhos de um profissional (e por consequência para receber o título de especialista na área), porém é de responsabilidade do usuário interessado definir como a restrição temporal deve afetar na recuperação da informação.

1.2 Objetivo Geral

O objetivo principal desse trabalho é propor e implementar um framework que seja capaz de inferir os principais tópicos de expertise dos diferentes pesquisadores cadastrados na Plataforma Lattes, dada sua produção, utilizando aprendizado não-supervisionado. Adicionalmente, o framework será testado em uma análise dos membros de um Programa de Pós-Graduação, mostrando as tendências e variações dos campos de pesquisa ao longo do tempo.

1.3 Objetivos Específicos

Para atingir o objetivo geral, os seguintes objetivos específicos foram definidos:

- i Identificar as informações dentro dos currículos que sejam relevantes para determinação da especialidade de cada pesquisador;
- ii Analisar os dados presentes nos currículos, selecionando os trechos de maior relevância para apuração da expertise;
- iii Propor um fluxo de tratamento para o conjunto de dados selecionados, de maneira que o texto seja processado para manter apenas as informações relevantes;
- iv Definir um modelo de aprendizado não-supervisionado que seja capaz de capturar os tópicos da expertise a partir dos documentos fornecidos, levando em consideração o fator tempo;
- v Propor um framework que utilize uma adaptação do modelo definido para inferir os tópicos da expertise das produções de cada pesquisador.

2 Fundamentação Teórica

Neste capítulo, é discutido o referencial teórico necessário para compreensão da proposta apresentada pelo presente trabalho. Inicialmente, a fonte de dados conhecida como *Currículo Lattes* é contextualizada, trazendo sua importância ao reunir currículos de pesquisadores e a relevância nas informações disponibilizadas, juntamente com a discussão da estruturação da informação dentro de cada um. Depois, são apresentados conceitos de recuperação de informação para extração da expertise, focando na técnica de modelagem por tópicos, mostrando também outras abordagens. A noção de tipos de aprendizado de máquina é resgatada, para concluir o capítulo com uma descrição do Latent Dirichlet Allocation (LDA) - uma modelagem de tópicos com aprendizado não supervisionado -, seguido do panorama sobre temporalidade na recuperação da expertise, um dos motivadores do trabalho.

2.1 Plataforma Lattes

A Plataforma Lattes representa a experiência do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na integração de bases de dados de Currículos, de Grupos de pesquisa e de Instituições em um único Sistema de Informações (CNPQ, 2018c). Dentro dessa plataforma, o módulo específico para tratar dos currículos é conhecido como *Currículo Lattes*, formado por um sistema web onde os estudantes e pesquisadores registram sua vida pregressa de maneira padronizada, sendo utilizado pela maioria das instituições de fomento, universidades e institutos de pesquisa em suas ações de planejamento, gestão e operacionalização de fomento do CNPq.

Desde a década de 80, o CNPQ já havia percebido a necessidade de utilização de um formulário padronizado para o registro das informações dos pesquisadores brasileiros, que permitiriam além de sua avaliação curricular, a seleção de consultores e especialistas, e geração de estatísticas (CNPQ, 2018b). Assim, em agosto de 1999, depois de muitas versões do formulário e de duas décadas de experiência nessa consolidação, juntamente com grupos universitários da UFSC, C.E.S.A.R e profissionais da Microsoft, a agência conseguiu lançar e padronizar o Currículo Lattes (CNPQ, 2018c). Desde então, o formulário vem sendo aprimorado para se adequar à realidade do cenário da pesquisa e às exigências do Ministério da Ciência, Tecnologia, Inovações e Comunicações.

O (CNPQ, 2018a) disponibiliza periodicamente um arquivo com todos os IDs dos pesquisadores cadastrados no Lattes. No momento da extração para este trabalho, a base de dados era datada de 08/07/2017, contando com 5.197.727 currículos cadastrados, distribuídos conforme as tabelas 1, 2 e 3. As informações contidas nessas tabelas foram

compiladas a partir do arquivo mencionado anteriormente.

| # | Área do conhecimento | Contagem | % do total |
|--------------|------------------------------------|------------------|-------------|
| 1 | <i>Não informado</i> | 1.860.668 | 35,8% |
| 2 | Educação | 258.930 | 5,0% |
| 3 | Administração | 241.377 | 4,6% |
| 4 | Direito | 219.096 | 4,2% |
| 5 | Medicina | 177.540 | 3,4% |
| 6 | Enfermagem | 163.494 | 3,1% |
| 7 | Ciência da Computação | 148.356 | 2,9% |
| 8 | Letras | 123.495 | 2,4% |
| 9 | Psicologia | 111.590 | 2,1% |
| 10 | Fisioterapia e Terapia Ocupacional | 88.514 | 1,7% |
| | Outras | 1.804.667 | 34,7% |
| Total | | 5.197.727 | 100% |

Tabela 1 – Distribuição dos currículos por área do conhecimento, conforme dados de julho de 2017

Fonte: Elaborado pelo autor

Os currículos estão distribuídos em 114 áreas do conhecimento, sendo que uma grande parcela (aproximadamente 35,8% do total) não possui essa informação cadastrada. Tais registros foram agrupados na opção *Não informado*. Entre as 115 opções, a Tabela 1 exhibe as 10 com maior concentração, agrupando as demais em *Outras*.

| # | Nacionalidade | Contagem | % do total |
|--------------|---------------|------------------|-------------|
| 1 | Brasileira | 5.129.564 | 98,69% |
| 2 | Colombiana | 9.424 | 0,18% |
| 3 | Peruana | 5.879 | 0,11% |
| 4 | Portuguesa | 5.033 | 0,10% |
| 5 | Argentina | 4.411 | 0,08% |
| 6 | Espanhola | 2.865 | 0,06% |
| 7 | Cubana | 2.628 | 0,05% |
| 8 | Americana | 2.585 | 0,05% |
| 9 | Moçambicana | 2.462 | 0,05% |
| 10 | Francesa | 2.424 | 0,05% |
| | Outras | 30.452 | 0,59% |
| Total | | 5.197.727 | 100% |

Tabela 2 – Distribuição dos currículos por nacionalidade, conforme dados de julho de 2017

Fonte: Elaborado pelo autor

Com relação à nacionalidade, os currículos estão distribuídos em 189 opções, sendo que na Tabela 2 são exibidos as 10 com a maior concentração. O destaque fica para o

Brasil, com a maioria absoluta, alcançando 98,69% do total. As demais nacionalidades foram agrupadas em *Outras*, que totalizaram 30.452 registros, ou 0,59% do total.

| # | Nível de formação | Contagem | % do total |
|--------------|---|------------------|-------------|
| 1 | Graduação | 1.508.017 | 29,01% |
| 2 | Ensino Médio (2o grau) | 1.450.678 | 27,91% |
| 3 | Especialização | 718.758 | 13,83% |
| 4 | Mestrado | 466.840 | 8,98% |
| 5 | <i>Não informado</i> | 350.675 | 6,75% |
| 6 | Doutorado | 271.633 | 5,23% |
| 7 | Ensino Profissional de nível técnico | 109.111 | 2,10% |
| 8 | Ensino Fundamental (1o grau) incompleto | 60.599 | 1,17% |
| 9 | Curso de curta duração | 56.238 | 1,08% |
| 10 | Ensino Fundamental (1o grau) | 55.038 | 1,06% |
| 11 | Especialização - Residência médica | 41.548 | 0,80% |
| 12 | Mestrado Profissional | 33.356 | 0,64% |
| 13 | Extensão universitária | 27.589 | 0,53% |
| 14 | MBA | 27.483 | 0,53% |
| 15 | Outros | 19.400 | 0,37% |
| 16 | Aperfeiçoamento | 764 | 0,01% |
| Total | | 5.197.727 | 100% |

Tabela 3 – Distribuição dos currículos por nível de formação, conforme dados de julho de 2017

Fonte: Elaborado pelo autor

Sobre o nível de formação, existem 14 opções específicas para cadastramento e outra genérica chamada de *Outros*. Além dessas 15 opções, foram encontrados registros sem o preenchimento dessa informação, que estão agrupados em *Não informado*. Os quantitativos de cada uma das 16 opções foram exibidos na Tabela 3. Levando em consideração os níveis mais altos de formação (Graduação, Especialização, Especialização - Residência médica, MBA, Mestrado, Mestrado Profissional e Doutorado), um total de 3.067.635 currículos é atingido, ou seja, 59,02% dos cadastrados. Como são níveis de formação que sugerem uma especialidade maior, a base do Currículo Lattes parece ser uma ótima fonte para busca de especialistas.

2.1.1 Estrutura de um currículo

Dentro do sistema da Plataforma Lattes, os currículos podem ser encontrados e acessados através das ferramentas de pesquisa disponibilizadas, ou através do seu link direto e individual, que é no formato <<http://lattes.cnpq.br/id>>, onde <id> é um número identificador do currículo, geralmente de 16 dígitos, como por exemplo <<https://lattes.cnpq.br/0055123456789102>>. Nesta página, praticamente todas as informações

cadastradas pelo pesquisador são exibidas em formato HTML, com um visual para facilitar a leitura e compreensão dos dados, conforme exemplificado pela Figura 1.

Prêmios e títulos

- 2012 CÁTEDRA IPEA-CAPE DE DESENVOLVIMENTO 2011-2012, IPEA-CAPE.
- 2010 CÁTEDRA IPEA-CAPE DE DESENVOLVIMENTO 2010-2011, IPEA-CAPE.
- 2007 3º Lugar no II Prêmio BRDE de Desenvolvimento/SC-2006-2007, BRDE-CORECON/SC.
- 2006 1º Lugar no I Prêmio BRDE de Desenvolvimento/SC 2005-2006, BRDE-CORECON/SC.

Produções

Produção bibliográfica

Artigos completos publicados em periódicos

Ordenar por

Ordem Cronológica

1.  RECKTENVALD, MARCELO ; MATTEI, LAURO ; PEREIRA, VILMAR ALVES . Avaliando o Programa Nacional de Assistência Estudantil (PNAES) sob a ótica das epistemologias. AVALIAÇÃO: REVISTA DA AVALIAÇÃO DA EDUCAÇÃO SUPERIOR, v. 23, p. 405-423, 2018.
2. MATTEI, L. Les conséquences socio-économiques du coup d'état parlementaire au Brésil. BRÉSIL(S) - SCIENCES HUMAINES ET SOCIALES, v. 1, p. 1-13, 2018.
3. MATTEI, L. A política agrária e os retrocessos do governo Temer. OKARA : GEOGRAFIA EM DEBATE (UFPB), v. 12, p. 293-307, 2018.
4. MATTEI, L.; SILVA, R. A. . O papel do microcrédito no ambiente local: evidências a partir do Programa Banco da Família em Santa Catarina. TEXTOS DE ECONOMIA, v. 21, p. 125-145, 2018.
5. MATTEI, L. Emprego e Desemprego em tempos de crise econômica e reforma trabalhista. REVISTA DO NÚCLEO DE ESTUDOS DE ECONOMIA CATARINENSE, v. 7, p. 2-10, 2018.

Figura 1 – Visualização de parte da página de um currículo cadastrado

Fonte: <<http://lattes.cnpq.br/5509228706749288>>

Apesar do formato HTML possuir uma estrutura e conferir uma semântica parcial aos dados, fica difícil proceder com a análise dos dados, uma vez que um título marcado com a tag <h1> pode se referir a qualquer grupo de informações. Para facilitar, a plataforma permite que cada currículo possa ser exportado e baixado em formato XML, dentro do próprio perfil dos pesquisadores. Estes arquivos XML, além de conter todas as informações cadastradas (diferente da versão HTML que mostra parcialmente), as marcações conferem uma semântica muito mais precisa aos dados. A formatação deste XML segue uma estrutura definida por um Document Type Definition (DTD), que pode ser conferido no anexo A. Parte da hierarquia das informações e alguns de seus atributos podem ser conferidos na listagem abaixo.

- DADOS-GERAIS (NOME-COMPLETO, NOME-EM-CITACOES-BIBLIOGRAFICA, DATA-NASCIMENTO, ...)
 - RESUMO-CV
 - FORMACAO-ACADEMICA-TITULACAO
 - * GRADUACAO (NIVEL, TITULO-DO-TRABALHO-DE-CONCLUSAO-DE-CURSO, NOME-INSTITUICAO, ANO-DE-INICIO, ANO-DE-CONCLUSAO,

- ...)
- * ESPECIALIZACAO
- * MESTRADO
- * DOUTORADO
- ATUACOES-PROFISSIONAIS
 - * ATUACAO-PROFISSIONAL (NOME-INSTITUICAO)
 - ATIVIDADES-DE-DIRECAO-E-ADMINISTRACAO
 - ATIVIDADES-DE-PESQUISA-E-DESENVOLVIMENTO
 - ATIVIDADES-DE-ENSINO
 - ATIVIDADES-DE-SERVICO-TECNICO-ESPECIALIZADO
- AREAS-DE-ATUACAO
- PREMIO-TITULOS
- PRODUCAO-BIBLIOGRAFICA
 - TRABALHOS-EM-EVENTOS
 - ARTIGOS-PUBLICADOS
 - * ARTIGO-PUBLICADO
 - DADOS-BASICOS-DO-ARTIGO (NATUREZA - COMPLETO | RESUMO | NAO_INFORMADO -, TITULO-DO-ARTIGO, ANO-DO-ARTIGO, PAIS-DE-PUBLICACAO, DOI, TITULO-DO-ARTIGO-INGLES, ...)
 - DETALHAMENTO-DO-ARTIGO (TITULO-DO-PERIODICO-OU-REVISTA, ISSN, VOLUME, ...)
 - AUTORES
 - PALAVRAS-CHAVE
 - AREAS-DO-CONHECIMENTO
 - SETORES-DE-ATIVIDADE
 - INFORMACOES-ADICIONAIS
 - LIVROS-E-CAPITULOS
 - TEXTOS-EM-JORNAIS-OU-REVISTAS
 - DEMAIS-TIPOS-DE-PRODUCAO-BIBLIOGRAFICA
 - ARTIGOS-ACEITOS-PARA-PUBLICACAO
- PRODUCAO-TECNICA
- OUTRA-PRODUCAO
- DADOS-COMPLEMENTARES

Como a hierarquia da informação dentro do documento é muito extensa, apenas parte dela foi relacionada na lista acima. Esse dados serão analisados mais adiante para compor a fonte de dados utilizada no trabalho. Nesse formato, os atributos possuem tipo (por vezes opções pré-definidas de escolha), e as tags possuem estrutura de relacionamento (hierarquia), quantidade e ocorrências também definidos dentro do DTD, conferindo uma precisão maior nos dados XML. Por exemplo, dentro da seção de *artigos publicados* (que fica dentro de *publicação bibliográfica*), cada artigo possui um subitem com os *dados básicos*, onde um de seus atributos é a *natureza* do artigo, que pode ser apenas uma das três opções: *completo*, *resumo ou não informado*. Esse tipo de definição, onde cada elemento possui atributos definidos (e por vezes seu possível conjunto de valores), ajuda na semântica e na determinação de quais informações são relevantes, porém não garante que eles estarão presentes no XML final (pois muitos são opcionais), nem com a informação correta.

2.2 Expertise Retrieval

A busca automática por especialistas é julgada desafiadora por vários autores. Motivados pelos recentes avanços nos sistemas de gerenciamento de dados e informações, muitas abordagens têm sido propostas e desenvolvidas pela academia e indústria, usando novas técnicas de recuperação da informação, data mining, descoberta do conhecimento, modelagem estatística, modelagem probabilística e redes complexas (LIN et al., 2017).

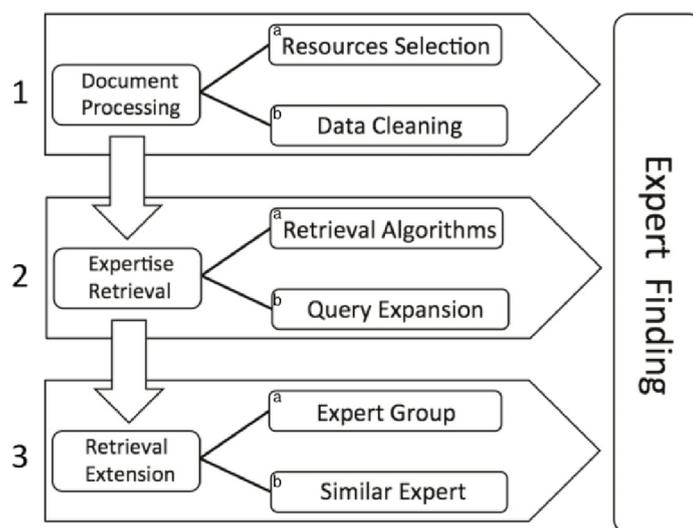


Figura 2 – Resumo dos procedimentos usados durante o *expert finding*

Fonte: Adaptado de Lin et al. (2017)

Na Figura 2 é possível verificar procedimentos básicos realizados durante o *expert finding*. Os autores definem que, em um primeiro momento, devem ser levantadas informações relevantes sobre a pessoa (1.a - *resource selection*), que podem ser encontradas em meta banco de dados (por exemplo: sistemas com informações de funcionários, bancos

de dados de universidades com as atividades de alunos e professores, etc), coleções de documentos (por exemplo: artigos e trabalhos acadêmicos publicados; documentos de trabalho, como e-mails, fóruns de perguntas, comentários e relatórios; páginas pessoais; etc) e redes de referência (geralmente criadas por pessoas que se agrupam e que possuem área de atuação semelhante, como grupos de pesquisa). Depois esses dados precisam ser tratados (*1.b - data cleaning*) antes de serem processados. A segunda etapa do processo é realmente processar os dados e fazer a recuperação da expertise (*2 - expertise retrieval*), onde algoritmos são aplicados. As abordagens usadas nessa etapa podem ser agrupadas conforme o tipo de modelo utilizado pelos algoritmos: modelos probabilísticos generativos (como os modelos de geração por candidatos e modelos de geração por tópicos), modelos de votação, modelos de grafos e um híbrido desses. Exemplos de modelos probabilísticos generativos por tópico estão descritos na Seção 2.2.1. Outros modelos são discutidos nas seções subsequentes. Por fim, o processo de recuperação da expertise é otimizado com uma etapa que aplica extensões nos algoritmos e tenta expandir (*3 - retrieval extension*) suas funcionalidades.

2.2.1 Modelagem por tópicos

A modelagem por tópicos é uma das técnicas utilizadas para extrair tópicos de maneira automática a partir de dados. Alguns esforços têm sido feitos nessa área utilizando métodos de clusterização, com objetivo de separar os documentos em grupos específicos, baseados em seu conteúdo semântico.

No trabalho de [Daud et al. \(2010\)](#), os seguintes modelos de tópicos, preexistentes na literatura, são citados como exemplos:

- *Probabilistic Latent Semantic Analysis* (PLSA): uma alternativa probabilística para dizer se um documento faz parte de um ou outro grupo. O PLSA é generativo apenas em nível de palavras, mas não de documentos (o que é considerado sua principal limitação);
- *Latent Dirichlet Allocation* (LDA): um modelo probabilístico que é generativo tanto a nível de palavras quanto de documentos. Isso possibilita o LDA prever os tópicos para novos documentos adicionados a um corpus prévio. Como este modelo é considerado o estado da arte para modelagem por tópicos, será discutido em detalhes, mais adiante;
- *Author-Topic* (AT): uma extensão do LDA que modela o interesse de autores com base em tópicos latentes.

Modelos probabilísticos de tópicos são um conjunto de algoritmos cujo objetivo é descobrir estruturas temáticas ocultas em grandes coleções de documentos ([FALEIROS](#);

LOPES, 2016). Quando várias palavras aparecem juntas com frequência (em documentos semanticamente relacionados), esse grupo pode ser considerado como um tópico.

Em um trabalho recente de um dos autores responsáveis pela criação do modelo LDA, são elencadas três ideias-chaves para a modelagem por tópicos (BLEI, 2012):

- Modelos de tópicos são algoritmos projetados para descobrir os temas principais que permeiam coleções de documentos largas e não estruturadas. Os modelos de tópicos podem organizar a coleção de acordo com os tópicos descobertos.
- São modelos que podem ser aplicados em coleções massivas de documentos.
- Esses modelos podem ser adaptados para muitos tipos de dados. Entre outras aplicações, eles têm sido utilizados para encontrar padrões em dados genéticos, imagens e redes sociais.

A modelagem por tópicos possui três vantagens principais com relação a outras abordagens de modelagem de documentos (ROSEN-ZVI et al., 2010):

1. Os tópicos são extraídos de uma maneira completamente não-supervisionada, não requerendo classificação de documentos ou pré-processamento;
2. Cada tópico é individualmente interpretável, provendo uma representação que pode ser compreendida pelo usuário;
3. Cada documento pode expressar múltiplos tópicos, capturando as combinações que surgem no texto.

2.2.2 Modelos com grafos de relacionamento

Para resolver o problema de localização de especialistas, alguns autores enxergam a problemática da relação entre documentos, tópicos e pessoas como um modelo de grafos, considerando diversos aspectos das informações disponíveis. Existem diferentes formas de modelar a estrutura, sendo as duas mais usuais (i) documentos e candidatos como nodos e suas associações como arestas; e (ii) apenas candidatos como nodos e seus relacionamentos como arestas. Geralmente os algoritmos HITS, PageRank e propagação *random walk* são utilizados em modelos de grafos (LIN et al., 2017). A ideia da classificação baseada em algoritmos de grafos é uma forma de decidir a importância de um nodo individual, levando em consideração a informação global, calculada recursivamente no grafo inteiro, ao invés de considerar apenas informações locais (específicas e individuais) (SORNIL, 2010).

No trabalho de Dom et al. (2003), a comunicação através de emails pode ser vista como um grafo direcionado, onde os nodos são pessoas, as arestas direcionadas os envios das mensagens, implicando, dentro do modelo, graus de especialidade e conhecimento em

quem a enviou. De maneira parecida, [Campbell et al. \(2003\)](#) considera quanto mais emails sobre um tópico um destinatário recebe, mais especialista ele é no tópico. Os tópicos são gerados com técnicas não supervisionadas de agrupamento em conjunto com palavras-chave informadas pelos próprios usuários.

2.2.3 Modelo de votação

Um modelo proposto por [Macdonald \(2009\)](#) é o Modelo de Votação. Neste caso, a busca por especialista é considerada como um processo de votação. Usando uma lista classificada dos documentos retornados a partir de um parâmetro de consulta Q , é proposto que a classificação dos candidatos pode ser modelada como um processo de votação, usando a classificação dos documentos e a lista de documentos do perfil de cada candidato. Isto vem de duas observações: (i) candidatos que escrevem muito sobre um tópico de interesse provavelmente podem ser considerados especialistas na área e (ii) quanto mais documentos em seu perfil estiverem relacionados à consulta Q , mais forte é a probabilidade de experiência. O problema principal consiste em como agregar os votos para criar a classificação final. Neste caso, são usadas técnicas de fusão de dados para combinar múltiplas classificações em uma única. Todas as vezes que um documento é recuperado de um perfil por um sistema de IR (baseado na consulta Q), um voto implícito é adicionado para que ele seja elevado na classificação combinada, indicando experiência relevante à consulta Q .

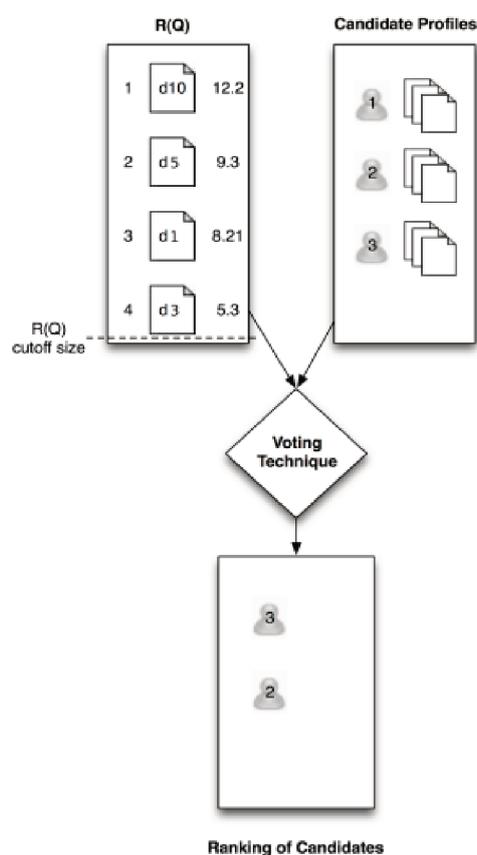


Figura 3 – Estrutura do modelo de votação

Fonte: Macdonald (2009)

Na Figura 3 é exemplificado como o modelo funciona: dada uma consulta Q , os documentos retornados de forma classificada a partir de sua aplicação são representados por $R(Q)$, que, juntamente com todos os documentos de cada perfil, são submetidos às técnicas de votação, responsáveis por calcular e fornecer a classificação dos candidatos com base na consulta Q . Dentro das técnicas de votação é onde ocorre a fusão de dados.

Esse modelo é baseado em um tipo de classificação muito antiga, socialmente usado por pessoas organizadas em grupos. Um exemplo é o sistema de escolha para cargos dentro de companhias ou na política: um grupo de pessoas consideradas aptas votam nos candidatos que julgam mais adequados para exercer a função. Diferentes tipos de sistemas de pesos e votação podem ser aplicados (sistemas de único ganhador, múltiplos ganhadores, monótonos, com prevenção de manipulação, neutros, entre outros).

Como as técnicas de votação aplicadas podem ser as mais diversas, fica fácil aplicar este modelo para outros cenários, com documentos obtidos através de sistemas de Perguntas & Respostas - onde as respostas recebem classificação pelos próprios usuários do sistema -, em documentos obtidos de redes sociais - a partir dos likes e comentários - ou sistemas colaborativos de equipes.

2.2.4 Outros modelos

Para nível de conhecimento, ainda existem outros modelos e abordagens discutidos na literatura:

- Para [Lin et al. \(2017\)](#), dentro da classe dos modelos probabilísticos ainda há o discriminativo (além dos generativos por tópico, já citados);
- No caso de [Zhu et al. \(2009\)](#), há uma discussão onde os processos de ER são sensíveis a várias características do documento, incluindo sua estrutura interna, a autoridade e múltiplos níveis de associação entre especialistas e tópicos;
- [Balog et al. \(2009\)](#) introduz um método de proximidade, onde o nível de dependência é medido pela distância entre os termos de consulta e a posição do candidato dentro do documento

2.3 Tipos de aprendizado

Dentro do ramo da Inteligência Artificial, quando um algoritmo aprende diferentes tipos de informação a partir dos dados que está analisando, ele é classificado como uma abordagem de Aprendizado de Máquina. Os algoritmos dentro dessa abordagem ainda podem ser subdivididos em algoritmos de aprendizado *supervisionado*, *não supervisionado* ou ainda um híbrido de ambos, os *semi-supervisionado* ([TALABIS et al., 2015](#)).

Algoritmos de aprendizado supervisionado usam dados rotulados anteriormente (geralmente chamado de *conjunto de treinamento*) para inferência do aprendizado, que é usado como base para classificação de novos dados que ainda não foram rotulados. As técnicas mais conhecidas são de *regressão linear* e *classificação* (k-vizinhos mais próximos, redes neurais, análise discriminante linear, entre outros).

No caso dos algoritmos de aprendizado não supervisionado, por não existir um conjunto de treinamento, toda a informação é obtida dos dados sendo analisados. Os dados são classificados a partir da descoberta da estrutura entre os elementos comuns presentes, sendo as principais técnicas o *Agrupamento (Clustering)* e *Análise dos Componentes Principais* ([TALABIS et al., 2015](#)).

2.4 Recuperação da expertise com o Latent Dirichlet Allocation: modelagem por tópicos com aprendizado não supervisionado

Como a ideia deste trabalho é utilizar uma abordagem não supervisionada para a recuperação da expertise e como muitos autores consideram o Latent Dirichlet Allocation (LDA) o estado da arte para modelagem por tópicos e pela sua importância, ele receberá

uma seção dedicada a sua explicação. O LDA é usado como base para vários outros algoritmos, sendo uma de suas extensões o guia para este trabalho.

Um modelo generativo de documentos é baseado nas regras probabilísticas de amostragem que descrevem como palavras em documentos podem ser geradas com base em variáveis latentes (aleatórias). Quando um modelo generativo é aplicado, o objetivo é encontrar o melhor conjunto de variáveis que podem explicar os dados observados (isto é, explicar as palavras observadas e suas relações [o porquê ocorreram daquela forma] nos documentos), assumindo que o modelo gerou os dados (STEYVERS; GRIFFITHS, 2007).

O LDA teve sua origem no probabilistic Latent Semantic Indexing (pLSI), proposto por Hofmann em 1999, que por sua vez, baseou-se em um modelo prévio, o Latent Semantic Indexing (LSI). Segundo a explicação de FALEIROS e LOPES (2016), no modelo pLSI, cada palavra em um documento é amostrada de uma variável aleatória que representa um tópico. Assim, cada palavra em um documento é gerada por um tópico, e cada documento possui palavras geradas por diferentes tópicos. Isso faz com que um documento possua diferentes proporções de tópicos. Apesar do pLSI ser um modelo probabilístico, ele não é um modelo gerador de documentos completo pois não provê um modelo probabilístico no nível dos documentos. Ou seja, apesar das palavras serem geradas por variáveis aleatórias obedecendo uma distribuição multinomial, os documentos são apenas *sacos de palavras*. O modelo pLSI foi estendido para o modelo LDA, que é um modelo bayesiano completo e se baseia na geração dos tópicos como distribuições de Dirichlet. Em comparação ao pLSI, o LDA descreve um modelo capaz de classificar documentos não conhecidos (documentos que não foram utilizados no treinamento), e utilizar informações *a priori*.

Seguindo a mesma explicação, o LDA não é um algoritmo com descrições sequenciais de instruções para encontrar tópicos dada uma coleção de documentos: é um modelo probabilístico no qual é descrito como os documentos são gerados. Nesse modelo, as variáveis observáveis são os termos de cada documento e as variáveis não observáveis são as distribuições de tópicos. Os parâmetros das distribuições de tópicos, conhecidos como hiper-parâmetros, são dados *a priori* no modelo.

Em resumo, o LDA expõe um método que, a partir de probabilidades previamente definidas para um conjunto de tópicos, um documento pode ser gerado da seguinte maneira: um tópico é escolhido levando em conta sua probabilidade, depois, uma palavra é escolhida seguindo as probabilidades de palavras dentro do tópico em questão (para cada tópico a lista de palavras e suas probabilidades são diferentes). O que se deseja fazer é justamente o processo inverso: dado um conjunto de documentos, descobrir quais tópicos estavam envolvidos, quais suas probabilidades e qual a composição interna de cada um (bem como probabilidade das palavras). Os algoritmos que fazem o processo inverso são não supervisionados.

Na imagem 4, pode ser verificada a representação visual do LDA. Os itens abaixo descrevem a notação utilizada.

- K - Número de tópicos
- n - Número de palavras do vocabulário
- m - Número de documentos
- n_{d_j} - Número de palavras em um documento d_j , onde $1 \leq j \leq m$
- θ - Distribuição de tópicos por documento
- ϕ - Distribuição dos tópicos sobre as palavras de todo o vocabulário
- θ_j - Vetor com a proporção dos tópicos para o documento d_j , onde $1 \leq j \leq m$
- ϕ_k - Vetor com a proporção das palavras do vocabulário para o tópico k , onde $1 \leq k \leq K$
- α - Priore da distribuição de Dirichlet, relacionada a distribuição documento-termo
- β - Priore da distribuição de Dirichlet, relacionada a distribuição tópico-palavra
- w_i - i -ésima palavra do vocabulário, onde $1 \leq i \leq n$
- $w_{j,i}$ - palavra w_i observada no documento d_j , onde $1 \leq j \leq m$ e $1 \leq i \leq n$.
- $z_{j,i}$ - Distribuição de tópicos associado a palavra $w_{j,i}$ no documento d_j , onde $1 \leq j \leq m$ e $1 \leq i \leq n$.

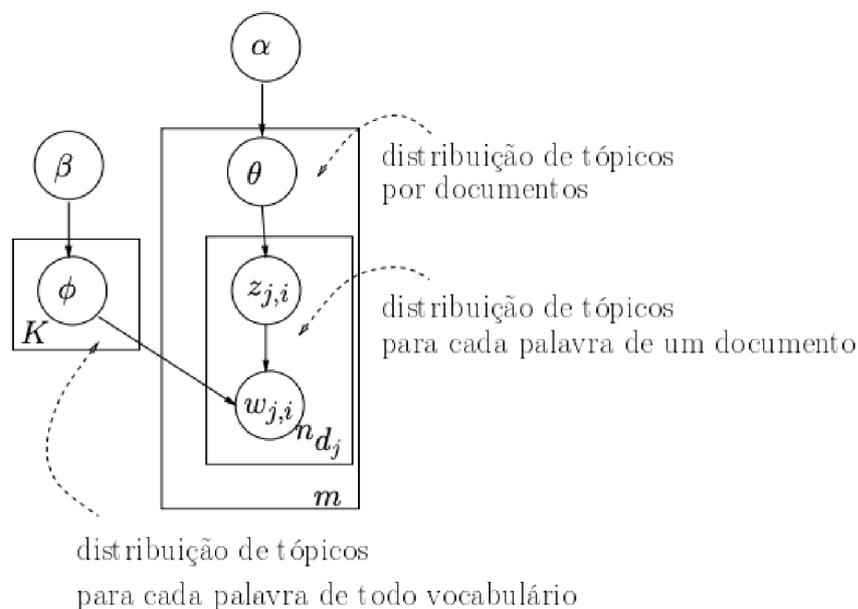


Figura 4 – Representação visual do LDA

Fonte: (FALEIROS; LOPES, 2016)

Utilizando a Figura 4 para esclarecer a representação do modelo, percebe-se que no nível de toda a coleção de documentos estão os hiper-parâmetros α e β . De forma simples, sem o formalismo matemático, pode-se dar uma interpretação para esses parâmetros. Um alto valor de α significa que cada documento provavelmente conterá uma maior mistura de tópicos. Um valor baixo para α indica maior probabilidade dos documentos conterem mistura de poucos tópicos, fazendo uma maior concentração em poucos tópicos. Da mesma forma, um valor alto para β significa que cada tópico terá alta probabilidade de conter misturas de várias palavras. Enquanto que um valor baixo para β indica que o tópico será formado por poucas palavras.

Para fazer a inferência do LDA (o processo inverso comentado anteriormente), muitos pesquisadores fazem uso do amostrador de Gibbs. O amostrador de Gibbs é um caso especial da simulação de Monte Carlo em Cadeia de Markov. Métodos de Monte Carlo em Cadeia de Markov podem emular distribuições de probabilidades com alta dimensionalidade por meio do comportamento estacionário da cadeia de Markov (FALEIROS; LOPES, 2016).

Levando em consideração a aplicação do LDA sobre os currículos de pesquisadores, entende-se que os tópicos gerados serão relacionados com os tópicos de expertise de cada um.

2.5 Temporalidade na identificação da especialidade

A especialidade das pessoas evolui com o tempo e, segundo Fang e Godavarthy (2014), apesar do grande trabalho desenvolvido nos últimos anos sobre recuperação da expertise, pouco tem sido estudado sobre o fator dinâmico inerente. Ainda segundo o autor, vários fatores podem afetar a dinâmica da especialidade individual, como tecnologias novas e emergentes tornando as existentes obsoletas, levando conseqüentemente a uma adaptação de habilidades e especialidades.

Conforme discutido na motivação deste trabalho, desconsiderar o fator tempo pode levar a resultados inadequados. Ao procurar revisores para um periódico, por exemplo, os organizadores podem estar interessados em especialistas que estejam produzindo nas áreas de interesse nos últimos anos. Caso a análise seja feita de todo o histórico do especialista, pode-se chegar a conclusão que é especialista no tópico T , mas na realidade, é um ramo não estudado há mais de 10 anos. Na Figura 5, é exemplificada a mudança gradativa (em diferentes pontos no tempo) dos tópicos de uma pessoa. Entre o ponto t_4 e t_9 , há um leve fortalecimento de ambos os ramos de pesquisa, mas em t_{10} há uma mudança de foco, enfraquecendo o ramo da esquerda e fortalecendo o ramo da direita. De maneira análoga, na Figura 6, é exibida a intensidade de cada ramo de pesquisa de um pesquisador. Entre 2002 e 2012, é perceptível uma convergência no foco para o ramo central, que no início da análise, em 1998, não era proeminente. Como estas são mudanças naturais na

carreira de um pesquisador, entende-se que o fator tempo deve ser considerado nas análises. Alguns modelos que abordam recuperação da expertise e temporalidade serão discutidos no Capítulo 3.

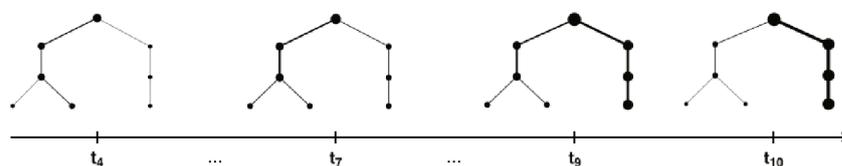


Figura 5 – Mudança dos tópicos de uma pessoa ao longo do tempo

Fonte: (RYBAK; BALOG; NØRVÅG, 2014)

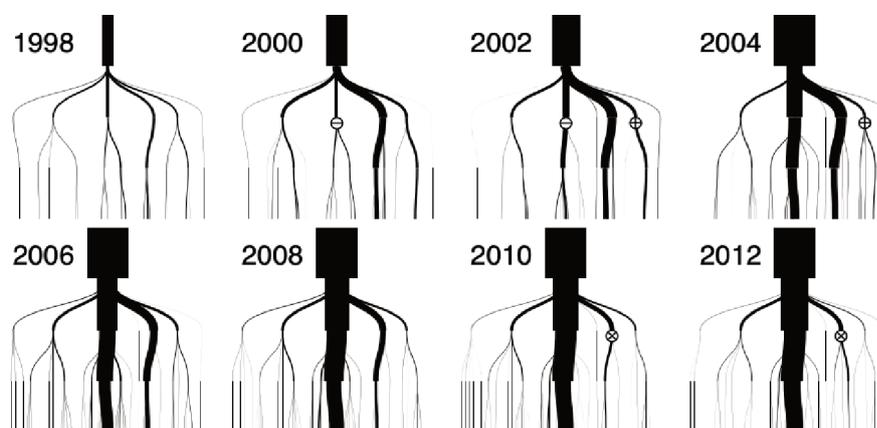


Figura 6 – Análise temporal da expertise de uma pessoa

Fonte: (RYBAK; BALOG; NØRVÅG, 2014)

3 Trabalhos Relacionados

Anteriormente, foram discutidos vários conceitos importantes para este trabalho. Neste capítulo, são discutidos modelos que consideram vários desses aspectos, mas que necessariamente integram a temporalidade na determinação da especialidade de pesquisadores.

3.1 Modelo Hierárquico

No trabalho de Rybak, Balog e Nørsvåg (2014) é criado um Modelo Hierárquico, definido através de uma árvore T_a de tópicos, onde os nodos filhos representam subtópicos de seus nodos pais, sendo que o peso de cada nodo identifica a expertise da pessoa naquele tópico. Inicialmente são calculados apenas os pesos dos nodos folhas e os demais de maneira *bottom-up*, onde o peso de um nodo intermediário é sempre a soma dos pesos de seus descendentes diretos. A taxonomia dos tópicos (isto é, a estrutura da árvore) deve ser previamente fornecida.

Para estimar o peso dos nodos folhas c_i , é utilizado um conjunto de documentos D_a que estão relacionados com um ou mais desses nodos. Para cada documento $d \in D_a$ a probabilidade $P(c|d)$ é definida para expressar o *quanto* o documento d pertence ao tópico c . A maioria dos documentos possui apenas um tópico relacionado, neste caso, $P(c|d) = 1$. Quando existem mais tópicos, esse peso é dividido igualmente entre todos, logo $P(c|d) = \frac{1}{|d_c|}$. Caso o tópico c não esteja relacionado com o documento, $P(c|d) = 0$. O somatório de todas as probabilidades dos documentos relacionados ao nodo c_i é considerada a expertise da pessoa a no tópico representado pelo nodo. Formalmente, o peso w_i do nodo folha i é dado por:

$$w_i = \sum_{d \in D_a} P(c_i|d)P(d)$$

Neste momento, os autores consideram a adição da probabilidade $P(d)$, que indica a importância do documento. Esse fator é utilizado fator para, por exemplo, dar importância maior aos artigos publicados em periódicos mais conceituados. Na elaboração do seu trabalho, os autores usam sempre $P(d) = 1$ argumentando que este fator será analisado em trabalhos futuros.

Para fazer o *temporal expertise profile*, a árvore T_a é construída diversas vezes ao longo do tempo, formando snapshots da forma $\tau_a = \{T_a^{t_1}, T_a^{t_2}, \dots, T_a^{t_m}\}$. O intervalo de tempo pode ser constante ou não. Neste ponto, o peso do nodo $w_i^{t_j}$ do snapshot t_j

é definido em duas partes: (i) expertise adquirida através dos documentos criados no intervalo $(t_{j-1}, t_j]$ e (ii) expertise carregada do passado, levando em consideração uma função de decaimento dando cada vez menos importância para documentos criados em snapshots distantes.

A caracterização de mudanças é feita analisando o peso dos nodos considerados como *foco* entre dois snapshots. Um nodo foco é aquele que acumula um peso considerável dentro de um snapshot. Caso a mudança entre $w_i^{t_j}$ e $w_i^{t_{j+1}}$ seja considerável, a posição dentro da árvore do nodo i é avaliada para dizer se está ocorrendo uma alteração a nível de campo de pesquisa (caso seja em nodos mais altos, como na Figura 8) ou a nível de tópico (nodos mais baixos, como na Figura 8).

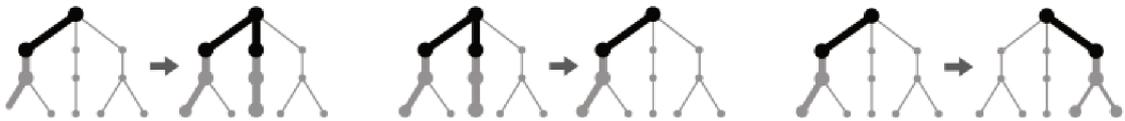


Figura 7 – Exemplo de mudança de campos de pesquisa, com alteração em nodos mais altos

Fonte: (RYBAK; BALOG; NØRVÅG, 2014)

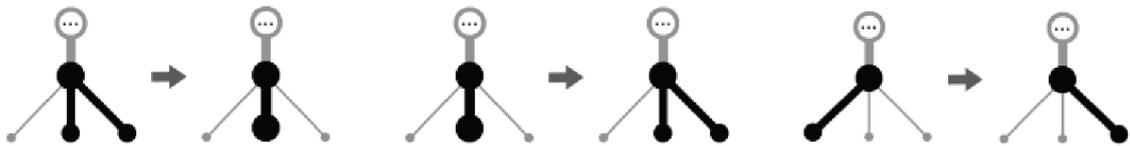


Figura 8 – Exemplo de troca de tópicos de interesse, dentro de um mesmo campo de pesquisa.

Fonte: (RYBAK; BALOG; NØRVÅG, 2014)

3.2 Modelo Temporal-Author-Topic

Na modelagem de interesses temporais de autores com tópicos, ou Temporal-Author-Topic (TAT), que é uma extensão do LDA, definidos por Daud et al. (2009), com mais detalhes em Daud (2012), um documento é visto como uma composição de palavras, onde cada uma é associada com o ano de publicação do trabalho e com seus respectivos autores. Uma coleção D pode ser escrita da forma $D = \{(w_1, a_1, y_d), (w_2, a_2, y_d), \dots, (w_d, a_d, y_d)\}$, sendo w_d um vetor de palavras de um vocabulário de tamanho V , a_d o vetor de autores, y_d o ano de publicação do trabalho d .

No modelo proposto, cada autor (do conjunto A de autores) do documento d é associado com uma distribuição multinomial θ_a sobre tópicos, e cada tópico é associado a

uma distribuição multinomial Φ_z sobre palavras e uma distribuição multinomial Ψ_z com o *timestamp* para cada palavra do tópico em questão. As distribuições anteriores são usadas, respectivamente, como os hyper-parâmetros α , β , γ do LDA.

Com o processo, os autores conseguem descobrir e classificar (conforme o valor da probabilidade) autores em áreas de pesquisa específicas, com base nos tópicos latentes de diferentes anos. Além disso, a abordagem do TAT permite determinar dinamicamente correlações entre autores em diferentes anos, calculando a dissimilaridade entre eles. Quando menor a dissimilaridade entre autores, maior a correlação entre eles. Vale ressaltar que não são necessariamente co-autores, mas autores que tendem a produzir tópicos semelhantes. Como o trabalho utiliza probabilidades em cada ano, para cada palavra dentro de cada tópico, é possível determinar os interesses de cada autor ao longo do tempo, dizendo qual foi seu foco de publicação em cada período.

A representação visual do modelo pode ser verificada na Figura 9.

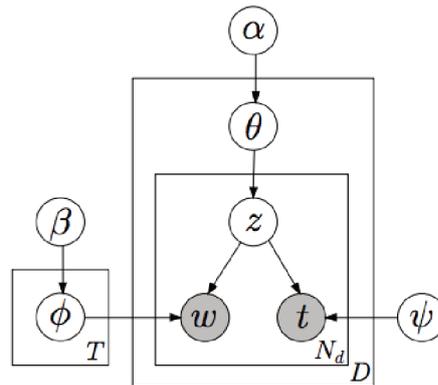


Figura 9 – Representação visual do TAT

Fonte: (DAUD, 2012)

3.3 Modelo Temporal-Expert-Topic

Em seu trabalho posterior, Daud et al. (2010) aprimora a abordagem TAT, modelando o *temporal expert finding* através de tópicos temporais generalizados, chamado de *Temporal-Expert-Topic* (TET). Nessa nova abordagem, são modelados simultaneamente a influência das conferências onde os trabalhos são publicados e as informações temporais. Aqui, um documento é um *subgrupo* e usualmente possui poucas palavras relacionadas (como o título e os autores), enquanto uma conferência é denotada como *grupo*, que usualmente possui vários trabalhos relacionados a um determinado tópico, o que leva a agrupar várias palavras semanticamente relacionadas. Os autores levam em consideração que um evento recebe todos os anos várias submissões de trabalhos escritos por diferentes autores, sendo que cada um contém um conjunto de palavras, no título e no nome dos autores, que cobrem a maior parte das subáreas de pesquisa dos eventos e dos autores.

Uma conferência c é denotada por um vetor de N_c palavras baseadas nos trabalhos aceitos para um ano y específico, um autor r com base em seus trabalhos aceitos, sendo que o *temporal expert finding* é formulado como: dado um evento c com N_c palavras possuindo um *stamp* y , descubra as pessoas mais qualificadas em um domínio específico para diferentes anos.

O TET consiste em duas etapas: (i) é utilizada a abordagem *Semantics and Temporal Information based Maven Search* (STMS) para calcular as matrizes de contagem $W \times Z$ (palavra por tópico), $Y \times Z$ (ano por tópico) e $Z \times R$ (ano por autor) e (ii) uma derivação do Teorema de Bayes para determinação dos experts relacionados aos tópicos por ano. O processo é semelhante ao TAT. Aqui, cada autor de um conjunto K de autores de uma conferência é considerado como responsável pela geração de alguns tópicos latentes dessa conferência. Por sua vez, esses tópicos são responsáveis por gerar as palavras e os *timestamps* para a conferência. Formalmente, cada autor do conjunto K de um evento c é associado com uma distribuição multinomial θ_r sobre tópicos e cada tópico é associado com uma distribuição Φ_z sobre palavras e uma distribuição multinomial Ψ_z com o ano para cada palavra de um evento do tópico em questão. Portanto, θ_r , Φ_z e Ψ_z possuem prior Dirichlet simétrico com os hiper-parâmetros α , β , γ , respectivamente.

O processo generativo é como segue:

1. Para cada autor $r = 1, \dots, K$ do evento c
 - a) Escolha θ_r a partir do *Dirichlet*(α)
2. Para cada tópico $z = 1, \dots, T$
 - a) Escolha Φ_z a partir do *Dirichlet*(β)
 - b) Escolha Ψ_z a partir do *Dirichlet*(γ)
3. Para cada palavra $w = 1, \dots, N_c$ do evento c
 - a) Escolha uniformemente um autor r entre todos os autores de a_c
 - b) Escolha um tópico z da *multinomial*(θ_r) condicionada em r
 - c) Escolha uma palavra w da *multinomial*(Φ_z) condicionada em z
 - d) Escolha um ano y associado com a palavra w da *multinomial*(Ψ_z) condicionada em z .

O amostrador de Gibbs é utilizado para estimar os parâmetros. O relacionamento das camadas do modelo pode conferido na Figura 10.

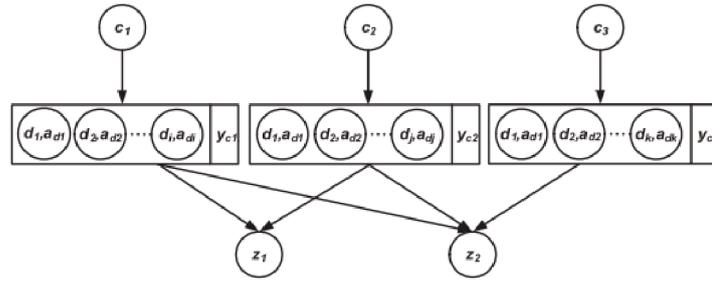


Figura 10 – Camadas presentes no TET.

Fonte: (DAUD et al., 2010)

O modelo é capaz de classificar probabilisticamente os experts em diferentes anos que estão relacionados com uma área de pesquisa específica. Como os interesses são dinâmicos, a cada ano a probabilidade dos autores pode ser alterada, o que leva a avaliação sobre seu interesse (e produção) no tópico em questão. Além disso, é possível calcular a dissimilaridade entre autores, o que traz a sua correlação de maneira dinâmica.

3.4 Modelo Random Walk

No trabalho de Li e Tang (2008), é utilizado o modelo *random walk*. Neste caso, uma rede é modelada em forma de grafo, sendo que diferentes tipos de objetos e diferentes tipos de relacionamento são modelados com diferentes tipos de nós e arestas. Essa abordagem é usada para contornar um dos problemas levantados pelos autores, onde a maioria dos modelos pré-existentes considera todos os objetos (trabalhos, conferência e autores) de maneira homogênea. Depois, para modelar a informação variando no tempo, conduzem uma propagação temporal *forward-and-backward* no modelo *random walk*.

A definição da rede heterogênea consiste de dois componentes: rede intra-tempo e rede inter-tempo. No caso de uma rede acadêmica, a rede intra-tempo é composta por outras 3 sub-redes:

- No centro está o grafo direcionado de citações de trabalhos $G_c = (V_c, E_c)$, onde:
 - V_c é o conjunto de todos trabalhos
 - E_c é o conjunto de arestas direcionadas de citações, onde $(d_i, d_j) \in E_c$ indica que o trabalho d_i citou o trabalho d_j .
- Para modelar o relacionamento de publicações entre autor-trabalho, é definido o grafo bipartido $G_{ac} = (V_a \cup V_c, E_{ac})$, onde:
 - V_a é o conjunto de autores
 - V_c é o conjunto de todos trabalhos

- E_{ac} é o conjunto de arestas de publicações, indicando quais autores são responsáveis por cada trabalho.
- O relacionamento entre trabalhos e locais de publicação (como conferências) é modelado com o grafo bipartido $G_{lc} = (V_l \cup V_c, E_{lc})$, onde:
 - V_l é o conjunto de locais de publicação
 - V_c é o conjunto de todos trabalhos
 - E_{lc} é o conjunto de arestas que indica quais trabalhos foram publicados em cada local.

Os três grafos são combinados para formar um grafo heterogêneo, centrado pela rede de citações: $G = (V_c \cup V_a \cup V_l, E_c \cup E_{ac} \cup E_{lc})$. Arestas sem direção são transformadas em duas arestas direcionadas (uma de ida e outra de volta) entre origem e destino. Ainda é definido um grafo que descreve as probabilidades de transição entre diferentes tipos de nós, que formaliza o comportamento da *superfície de caminhada*.

São propostos dois métodos para incorporar a informação temporal ao modelo *random walk*:

- **Método de agregação:** A rede heterogênea G é separada em diferentes fatias de tempo $\{G_s\}$. Depois o *random walk* é executado nas diferentes janelas de tempo e combina os resultados do ranking de cada uma. Nesse método, os pesos das redes podem ser facilmente ajustados conforme o interesse pelo período específico. Por exemplo, caso haja interesse em objetos recentes, a janela de tempo referente ao período desejado pode ter peso mais alto, enquanto janelas de tempo anteriores recebem pesos menores. Uma função de decaimento é utilizada para controlar os scores de diferentes janelas de tempo.
- **Temporal random walk:** Aqui a intenção é integrar a variação do tempo diretamente no processo de caminhada dentro do grafo. Assumindo que existem os *timestamps* $\{1, 2, \dots, t\}$, G_s é definido como rede intra-tempo e $G_{(s-1)s}$ como rede inter-tempo, onde $G_{(s-1)s}$ é um grafo bipartido entre as redes intra-tempo vizinhas $G_{(s-1)s}$ e G_s . Vale notar que um objeto nesse modelo pode aparecer em mais de uma fatia de tempo. Por exemplo, um trabalho foi publicado na janela $(s-1)$, porém foi citado em um trabalho de (s) . Também são criadas probabilidades de transição entre os objetos de $(s-1)$ e (s) , λ_f *forward-transition* e λ_b *backward-transition*, que indica a probabilidade. A relação entre as probabilidades e as janelas de tempo pode ser verificada na Figura 11.

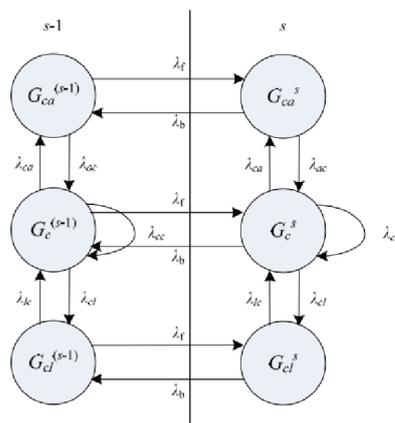


Figura 11 – Temporal random-walk
Fonte: Li e Tang (2008)

Nos experimentos desse trabalho houve interação humana na classificação inicial, pois os autores julgaram difícil encontrar um dataset confiável para os testes. Foram coletadas 7 consultas mais frequentes da rede Arnetminer (um serviço online para busca, indexação e mineração de grandes quantidades de dados da comunidade científica) e, para cada uma, os 30 trabalhos mais relevantes retornados. As mesmas consultas foram aplicadas também em outras duas redes similares, *Libra author search* e *Rexa authors search*. Esses documentos foram recuperados e unificados, sendo que os duplicados e com nomes inexistentes no Arnetminer foram removidos. Depois disso, 5 estudantes e 2 professores foram solicitados a julgar em uma escala de 4 pontos cada autor, trabalho e conferência. Uma especificação foi dada para nivelamento, como por exemplo, os trabalhos deveriam ser classificados conforme citações, relevância do conteúdo, ano de publicação e impacto para o campo de pesquisa.

3.5 Modelo Topics Over Time

Os autores Wang e McCallum (2006) apresentam um modelo de tópicos baseado no LDA, chamado de *Topics over Time* (TOT), que é capaz de capturar não apenas as mudanças de baixa dimensão na estrutura dos dados, mas também como elas mudam ao longo do tempo. Algumas das diferenças com modelos prévios de objetivos semelhantes são (i) utilizar o tempo na forma contínua (i.e. não é discretizado) e (ii) não fazer suposições de Markov sobre as transições de estado durante o tempo. O TOT parametriza uma distribuição contínua sobre o tempo associado com cada tópico, sendo estes tópicos os responsáveis por gerar tanto os *timestamps*, quanto as palavras.

Quando um forte padrão de coocorrência de palavras aparece por um breve período de tempo e depois desaparece, o TOT criará um tópico com uma estreita distribuição no tempo. Quando um padrão permanece consistente por um longo período, o modelo criará um tópico com uma ampla distribuição no tempo.

Na descrição do trabalho, uma distribuição Beta é utilizada sobre o período (que é normalizado em um intervalo de $[0 - 1]$). Para um conjunto de documentos, todas as palavras são associadas com o *timestamp* do documento.

Neste modelo, os tópicos e os seus significados são modelados como constantes ao longo do tempo. O TOT captura mudanças na ocorrência (e coocorrência condicionada no tempo) dos próprios tópicos, não mudanças nas distribuições das palavras de cada tópico. Na análise dos autores, isso pode ser por um lado útil e por outro perigoso. Um conjunto de documentos com fortes padrões de coocorrências, primeiro entre *pássaros* e *aerodinâmica*, depois entre *aerodinâmica* e *calor*, depois entre *calor* e *mecânica quântica* poderia levar a um único tópico que seguiu essa trajetória, levando a uma conclusão inapropriada que *pássaros* e *mecânica quântica* são versões do mesmo tópico deslocadas no tempo. Ainda segundo os autores, alternativamente, aplicando em vários documentos do campo da medicina (uma área que mudou drasticamente ao longo do tempo), fica fácil perceber as tendências nos tópicos que emergiram, deixaram de existir, ou ainda aqueles que ocorrem constantemente.

O TOT é visto como um processo generativo de *timestamps* e de palavras em documentos com *timestamp*, da forma como segue abaixo. A representação visual pode ser verificada na Figura 12 e a notação utilizada na Tabela 4.

1. Defina T distribuições multinomiais ϕ_z a partir do Dirichlet prior β , uma para cada tópico z ;
2. Para cada documento d , defina uma distribuição multinomial θ_d a partir do Dirichlet prior α ; então, para cada palavra w_{di} do documento d :
 - a) Defina um tópico z_{di} a partir da distribuição multinomial θ_d ;
 - b) Defina uma palavra w_{di} a partir da distribuição multinomial $\phi_{z_{di}}$;
 - c) Defina um *timestamp* t_{di} a partir de $Beta\psi_{z_{di}}$

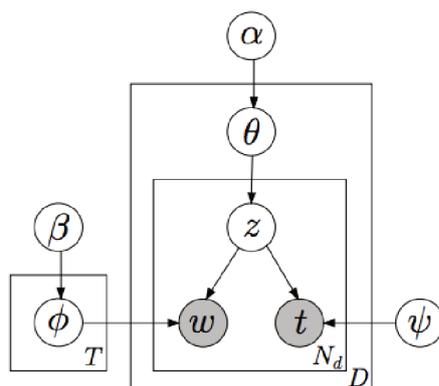


Figura 12 – Representação visual (no formato *plate notation*) do TOT, com o amostrador de Gibbs.

Fonte: (WANG; MCCALLUM, 2006)

| Símbolo | Descrição |
|------------|--|
| T | número de tópicos |
| D | número de documentos |
| V | número de palavras únicas (considerando todos os documentos) |
| N_d | quantidade total de palavras no documento d |
| θ_d | distribuição multinomial de tópicos, específica do documento d |
| ϕ_z | distribuição multinomial de palavras, específica do tópico z |
| ψ_z | distribuição beta do tempo, específica do tópico z |
| z_{di} | tópico associado com a i^{a} palavra do documento d |
| t_{di} | o <i>timestamp</i> associado com a i^{a} palavra do documento d |

Tabela 4 – Notação utilizada pelo TOT

Fonte: (WANG; MCCALLUM, 2006)

Vale lembrar que da mesma forma que o LDA, o TOT considera os documentos como *bag-of-words*, portanto a ordem das palavras pode ser negligenciada, sendo capturada apenas sua coocorrência.

Essa abordagem possui outras características, como ser capaz de dizer o ano do documento (de maneira muito aproximada) baseado em suas palavras.

4 O Framework Exper.te

Neste capítulo, é apresentado o framework Exper.te, proposto para determinação das especialidades de um pesquisador, levando em consideração o tempo e sua produção cadastrada no currículo da Plataforma Lattes, com base nos modelos analisados no capítulo anterior.

4.1 Visão Geral

Para cumprir com os objetivos do trabalho, o framework Exper.te foi projetado para receber como entrada um arquivo XML de um currículo, fazer o processamento das informações e, ao final, fornecer dados que sejam suficientes para determinação da especialidade de um pesquisador ao longo do tempo. O encadeamento dos passos do framework pode ser conferido na Figura 13 e sua discussão nas seções seguintes. Entre cada um dos passos, é indicado o tipo de informação que transita (saída do passo anterior e entrada para o próximo).

De maneira resumida, o Passo 1 é referente a leitura e análise dos arquivos texto (em formato XML) obtidos diretamente na Plataforma Lattes, fornecendo como saída um objeto XML para cada currículo (que permite o acesso direto aos elementos e atributos). No Passo 2, cada currículo é analisado e diferentes documentos são gerados, cada um com as informações relevantes (um documento para cada artigo publicado, por exemplo), fornecendo como saída um conjunto de documentos para cada pesquisador. No Passo 3, cada documento é traduzido para a Língua Portuguesa. No Passo 4, todas as informações irrelevantes são removidas (pontuação, palavras desnecessárias ou pequenas, números, entre outras) e palavras compostas são unificadas, disponibilizando para o próximo passo um conjunto de documentos contendo apenas palavras relevantes. No Passo 5, os algoritmos para identificação da expertise são parametrizados e executados, fornecendo as estatísticas necessárias para que o Passo 6 possa exibir os resultados em formas de gráfico ao usuário. Caso seja necessário, as distribuições das probabilidades também estarão disponíveis ao final do passo 6.



Figura 13 – Representação do fluxo do framework

Fonte: Elaborado pelo autor

4.2 Concepção do framework

Nesta seção são discutidas as informações mais detalhadas para a concepção do framework, juntamente com detalhes da implementação.

4.2.1 Fonte de dados, processamento de arquivos e geração dos documentos

Como descrito na Seção 2.1, a Plataforma Lattes é um grande repositório de currículos de pesquisadores, professores, alunos e outros profissionais, na sua maioria com

grau de formação elevado, o que sugere uma boa fonte de dados de especialistas. Além disso, o CNPQ possui anos de experiência no cadastro dessas informações, tendo padronizado o formulário de cadastro, que é disponibilizado em um formato semi-estruturado (apesar dos dados estarem disponíveis no formato XML, existem vários campos onde o texto é livre, sem marcação específica). Por esses motivos, somados ainda à falta de trabalhos que unem a busca de especialistas com essa base de dados tão relevante, ela foi escolhida para este trabalho.

O primeiro passo a ser realizado é a leitura dos arquivos correspondentes aos currículos e processamento no formato XML (Passo 1.a), bem como extração dos dados (Passo 1.b), que consistem basicamente na leitura dos arquivos texto, análise da estrutura e transformação em objetos XML (permitindo navegação e acesso aos dados). Em seguida, inicia-se a transformação do objeto XML, em um formato que possa ser utilizado pelo framework, gerando vários documentos (Passo 2) a partir da extração de informações relevantes. O XML é composto por vários elementos que correspondem à produção do pesquisador, incluindo artigos, projetos, orientações, atuações profissionais, formação acadêmica, entre outros. Cada ocorrência de elemento que corresponde a uma produção do pesquisador, com seus respectivos atributos, é usada na geração de documentos para o framework. Por exemplo, para cada artigo cadastrado, um documento individual será gerado a partir da transformação de um artigo, considerando seu título, nome do periódico, palavras chave e descrição adicional. Uma outra abordagem seria gerar um documento para o título, outro para o resumo e outro a concatenação de palavras chave. Um exemplo real é discutido adiante, onde dois elementos XML, presentes nos Códigos 1 e 2, são usados para gerar dois documentos listados no Código 3. Dessa maneira, um currículo vai gerar uma série de documentos, que nada mais são do que a transformação dos elementos do XML original em um novo formato de dados, onde os atributos relevantes são concatenados. Esse procedimento é usado para filtrar as informações dentro do currículo e também para montar os documentos usados na análise dos algoritmos de expertise, deixando-os mais completos (um documento gerado a partir de um artigo não será apenas o seu título, mas a concatenação do seu título, título do periódico, palavras-chave e informações adicionais).

```

1 <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="1364" ORDEM-IMPORTANCIA="22">
2   <DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="
   TRATAMENTO DE UM EFLUENTE MODELO TEXTIL VIA REACAO DE FENTON" ANO
   -DO-ARTIGO="2016" PAIS-DE-PUBLICACAO="" IDIOMA="Portugues" MEIO-
   DE-DIVULGACAO="MEIO_DIGITAL" HOME-PAGE-DO-TRABALHO="[doi
   :10.15628/holos.2016.2250]" FLAG-RELEVANCIA="NAO" DOI="10.15628/
   holos.2016.2250" TITULO-DO-ARTIGO-INGLES="TRATAMENTO DE UM
   EFLUENTE MODELO TEXTIL VIA REACAO DE FENTON" FLAG-DIVULGACAO-
   CIENTIFICA="NAO"/>
3   <DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="Holos (Natal
   . Online)" ISSN="18071600" VOLUME="4" FASCICULO="" SERIE=""
   PAGINA-INICIAL="25" PAGINA-FINAL="35" LOCAL-DE-PUBLICACAO=""/>
4   <AUTORES NOME-COMPLETO-DO-AUTOR="KOSLOWSKI, LUCIANO ANDRE DEITOS"
   NOME-PARA-CITACAO="KOSLOWSKI, LUCIANO ANDRE DEITOS" ORDEM-DE-
   AUTORIA="1" NRO-ID-CNPQ=""/>
5   <AUTORES NOME-COMPLETO-DO-AUTOR="LICODIEDOFF, SILVANA" NOME-PARA-
   CITACAO="LICODIEDOFF, SILVANA" ORDEM-DE-AUTORIA="2" NRO-ID-CNPQ="
   "/>
6   <AUTORES NOME-COMPLETO-DO-AUTOR="Humberto Gracher Riella" NOME-PARA-
   CITACAO="RIELLA, Humberto Gracher" ORDEM-DE-AUTORIA="3" NRO-ID-
   CNPQ=""/>
7   <PALAVRAS-CHAVE PALAVRA-CHAVE-1="reacao fenton" PALAVRA-CHAVE-2="
   efluente textil" PALAVRA-CHAVE-3="" PALAVRA-CHAVE-4="" PALAVRA-
   CHAVE-5="" PALAVRA-CHAVE-6=""/>
8   <AREAS-DO-CONHECIMENTO>
9     <AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="
       ENGENHARIAS" NOME-DA-AREA-DO-CONHECIMENTO="" NOME-DA-SUB-AREA
       -DO-CONHECIMENTO="" NOME-DA-ESPECIALIDADE=""/>
10  </AREAS-DO-CONHECIMENTO>
11  <SETORES-DE-ATIVIDADE SETOR-DE-ATIVIDADE-1="Pesquisa e
       desenvolvimento cientifico" SETOR-DE-ATIVIDADE-2="" SETOR-DE-
       ATIVIDADE-3=""/>
12  <INFORMACOES-ADICIONAIS DESCRICAO-INFORMACOES-ADICIONAIS="0 presente
       estudo teve como objetivo avaliar a aplicacao de um Processo
       Oxidativo Avancado no tratamento de efluentes texteis, denominado
       processo oxidativo de Fenton.
13  Nos estudos realizados, envolvendo a degradacao dos corantes Azul [...]
       />
14 </ARTIGO-PUBLICADO>

```

Código 1 – Exemplo de elemento que representa um artigo dentro do XML

Fonte: Elaborado pelo autor

```

1 <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="786" ORDEM-IMPORTANCIA="104">
2   <DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="O
   ESTADO DA ARTE NA RECICLAGEM DE LAMPADAS FLUORESCENTES NO BRASIL:
   PARTE 1" ANO-DO-ARTIGO="2008" PAIS-DE-PUBLICACAO="" IDIOMA=""
   Portugues" MEIO-DE-DIVULGACAO="IMPRESSO" HOME-PAGE-DO-TRABALHO=""
   FLAG-RELEVANCIA="NAO" DOI="" TITULO-DO-ARTIGO-INGLES="O ESTADO
   DA ARTE NA RECICLAGEM DE LAMPADAS FLUORESCENTES NO BRASIL: PARTE
   1" FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
3   <DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="Acta
   Ambiental Catarinense" ISSN="16777298" VOLUME="5" FASCICULO=""
   SERIE="1" PAGINA-INICIAL="43" PAGINA-FINAL="53" LOCAL-DE-
   PUBLICACAO="" />
4   <AUTORES NOME-COMPLETO-DO-AUTOR="Vera Lucia Mombach" NOME-PARA-
   CITACAO="Vera Lucia Mombach" ORDEM-DE-AUTORIA="1" NRO-ID-CNPQ="
   3464644628449312"/>
5   <AUTORES NOME-COMPLETO-DO-AUTOR="Humberto Gracher Riella" NOME-PARA
   -CITACAO="RIELLA, Humberto Gracher;Riella, H. G.;Riella, H.G.;
   FIORI JUNIOR, J.;RIELLA,HUMBERTO G.;RIELLA, HUMBERTO G.;GRACHER
   RIELLA, HUMBERTO;H.G.Riella" ORDEM-DE-AUTORIA="2" NRO-ID-CNPQ="
   0055080134892269"/>
6   <AUTORES NOME-COMPLETO-DO-AUTOR="Nivaldo Cabral Kuhnen" NOME-PARA-
   CITACAO="KUHNNEN, Nivaldo Cabral" ORDEM-DE-AUTORIA="3" NRO-ID-CNPQ
   ="7996380815454623"/>
7   <PALAVRAS-CHAVE PALAVRA-CHAVE-1="lampadas fluorescentes" PALAVRA-
   CHAVE-2="Reciclagem" PALAVRA-CHAVE-3="" PALAVRA-CHAVE-4=""
   PALAVRA-CHAVE-5="" PALAVRA-CHAVE-6="" />
8   <AREAS-DO-CONHECIMENTO>
9     <AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="
       ENGENHARIAS" NOME-DA-AREA-DO-CONHECIMENTO="" NOME-DA-SUB-AREA
       -DO-CONHECIMENTO="" NOME-DA-ESPECIALIDADE="" />
10  </AREAS-DO-CONHECIMENTO>
11  <SETORES-DE-ATIVIDADE SETOR-DE-ATIVIDADE-1="Industrias de
       Transformacao" SETOR-DE-ATIVIDADE-2="" SETOR-DE-ATIVIDADE-3="" />
12  <INFORMACOES-ADICIONAIS DESCRICAO-INFORMACOES-ADICIONAIS="As
       lampadas fluorescentes sao um artigo comum na sociedade atual. No
       entanto sao agentes de contaminacao ambiental quando descartadas
       inadequadamente. Algumas empresas tem se dedicado ao
       desenvolvimento de tecnologias de descontaminacao deste material
       [...]" />
13 </ARTIGO-PUBLICADO>

```

Código 2 – Outro elemento que representa um artigo, extraído do mesmo XML.

Fonte: Elaborado pelo autor

```

1 inicio;fim;documento
2 2016;2016;TRATAMENTO DE UM EFLUENTE MODELO TEXTIL VIA REACAO DE FENTON.
   Holos (Natal. Online). reacao fenton. efluente textil. O presente
   estudo teve como objetivo avaliar a aplicacao de um Processo
   Oxidativo Avancado no tratamento de efluentes texteis, denominado
   processo oxidativo de Fenton. Nos estudos realizados, envolvendo a
   degradacao dos corantes Azul Turquesa e Verde Malaquita, por meio de
   um planejamento fatorial, analises de absorbancia, DQO e pH, ficou
   evidenciada a elevada capacidade de degradacao da reacao de Fenton.
3
4
5 2008;2008;O ESTADO DA ARTE NA RECICLAGEM DE LAMPADAS FLUORESCENTES NO
   BRASIL: PARTE 1. Acta Ambiental Catarinense. lampadas fluorescentes.
   Reciclagem. As lampadas fluorescentes sao um artigo comum na
   sociedade atual. No entanto sao agentes de contaminacao ambiental
   quando descartadas inadequadamente. Algumas empresas tem se dedicado
   ao desenvolvimento de tecnologias de descontaminacao deste material
   [...]

```

Código 3 – Documentos gerados a partir dos artigos apresentados nos Códigos 1 e 2, respectivamente.

Fonte: Elaborado pelo autor

A transformação dos documentos ocorreu da seguinte forma: com a ajuda do DTD que define a estrutura do XML do Currículo Lattes (disponível no Anexo A), os principais elementos e atributos foram extraídos e agrupados. Essa seleção foi feita de maneira empírica, a partir da análise de dados de currículos já preenchidos e do julgamento dos dados que são relevantes. Por exemplo: o atributo *NOME-GRANDE-AREA-CONHECIMENTO*, presente em vários tipos de produção (como nos artigos), está sempre (ou quase sempre) preenchido com o valor *ENGENHARIAS* para os pesquisadores das áreas da engenharia. Esta é uma informação que não agrega no momento do recuperação da expertise (por ser genérica), diferente das palavras-chave, que dizem muito sobre uma produção (informação mais específica). Os dados considerados relevantes para solidificação do modelo (i.e. que podem representar a expertise do pesquisador) foram relacionados no Apêndice A, no formato JSON. Essa estrutura é a base para a busca das informações dentro do XML. Nela foram adicionados registros com expressões XML Path Language (XPath) capazes de selecionar informações dentro do XML. Por exemplo, considerando que os artigos são importantes na determinação da expertise do pesquisador, foi verificado no DTD que a expressão XPath que recupera todos elementos que representam artigos é dada por `<./PRODUCAO-BIBLIOGRAFICA/ARTIGOS-PUBLICADOS/ARTIGO-PUBLICADO>`, sendo que o título de um artigo pode ser acessado por `<./DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO>` e o ano de publicação por `<./DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO>`. Essas são as informações que devem constar no JSON para identificar título e ano de cada artigo. Como existem outros atributos interessantes para esse tipo de produção, outras expressões XPath também foram adicionadas e podem ser conferidas no Código 4. Esse formato facilita a customização do framework caso a estrutura do Currículo Lattes sofra alteração (sendo necessário alterar apenas as expressões que recuperam as informações), ou ainda na troca da fonte de dados (fazendo novas expressões para busca em uma estrutura XML completamente diferente).

```
1 {
2   "descricao" : "Artigos publicados",
3   "tag" : "://PRODUCAO-BIBLIOGRAFICA/ARTIGOS-PUBLICADOS/ARTIGO-PUBLICADO",
4   "inicio" : ["/DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO"],
5   "fim" : ["/DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO"],
6   "atributos" : [
7     "/DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO",
8     "/DETALHAMENTO-DO-ARTIGO/@TITULO-DO-PERIODICO-OU-REVISTA",
9     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
10    "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
11    "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
12    "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
13    "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
14    "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
15    "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
16  ]
17 }
```

Código 4 – Definição de um item responsável por extrair os *artigos publicados* do XML de um currículo

Fonte: Elaborado pelo autor

Para manter a organização, cada registro dentro do JSON deve seguir a estrutura definida abaixo:

- **descricao:** uma descrição sobre o que se trata o item, utilizado para facilitar a visualização e depuração;
- **tag:** expressão XPath que seleciona um elemento (tag), ou vários, dentro do XML;
- **inicio:** uma lista com várias expressões XPath para selecionar o valor do ano de início do documento. A busca dessa XPath é feita a partir do elemento selecionado por **tag**. Quando houver mais de um item nessa lista, utiliza-se o valor da primeira expressão XPath que retornar um ano válido (pois os atributos podem estar vazios). Como este valor é obrigatório para o framework, quando não é encontrado um valor, o ano corrente é usado (para considerar o documento em algum momento e não simplesmente descartá-lo);
- **fim:** da mesma forma que **inicio**, porém utilizado para selecionar o ano final do documento (projetos, orientações, formação, entre outros, possuem duração maior que um ano);
- **atributos:** Uma lista com várias expressões XPath para elementos ou atributos, que serão o corpo dos documentos transformados. A aplicação desse XPath deve sempre retornar textos (isto é, caso um elemento seja selecionado, deve ser acrescentado ao final `"/text()`).

Com todos os registros inseridos no JSON, é possível aplicar a expressão XPath definida na chave **tag** de cada um sobre o XML completo, com o objetivo de extrair

informações e gerar os documentos desejados. Conforme discutido anteriormente, o Código 4 representa um registro responsável por selecionar todos os elementos referentes aos *artigos publicados*. A aplicação da XPath `<./PRODUCAO-BIBLIOGRAFICA/ARTIGOS-PUBLICADOS/ARTIGO-PUBLICADO>` em um currículo¹, retorna um conjunto elementos, que serão chamados individualmente de *e*. Dois exemplos dos elementos XML retornados por esta expressão XPath podem ser conferidos nos Códigos 1 e 2.

Para cada elemento *e* retornado, é definido o ano da produção em questão, que tem seu início e fim definidos pelos atributos `inicio` e `fim`. Para os dois artigos dos Códigos 1 e 2, `inicio` e `fim` são definidos com a expressão `<./DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO>`, sendo que sua aplicação informa que o período do primeiro artigo será de 2016–2016 e do segundo 2008–2008. Como as duas XPaths são iguais, o início e fim da produção de artigo serão iguais. Esse valor só será diferente para os trabalhos que duram mais de um ano, como projetos e orientações de mestrado ou doutorado. No caso do `fim`, em alguns registros existem duas XPaths, como é o caso da *Formação acadêmica - Mestrado*: o final do mestrado pode ser definido pelo atributo `<./@ANO-DE-OBTENCAO-DO-TITULO>` ou `<./@ANO-DE-CONCLUSAO>`, sendo utilizado o primeiro que retornar valor.

Ainda sobre *e*, são aplicadas as XPaths definidas na chave `atributos`, onde cada valor retornado é concatenado em um único documento. No caso dos artigos, serão concatenados os valores dos atributos: `<./DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO>`, `<./DETALHAMENTO-DO-ARTIGO/@TITULO-DO-PERIODICO-OU-REVISTA>`, palavras chave de `<./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1>` a `<./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6>` e `<./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS>` (geralmente onde consta o resumo). No Código 3 são mostrados os dois documentos gerados para os artigos discutidos anteriormente. Neste momento, o interesse é apenas gerar os documentos. Em um segundo momento, eles serão tratados e filtrados, para depois serem utilizados na recuperação da expertise.

Além dos artigos, vários outros dados são extraídos do currículo durante a geração dos documentos. Na lista abaixo, estão descritos, de maneira resumida, os grandes grupos de informações. A lista completa, com os detalhes de cada elemento e atributo extraídos estão relacionados no Apêndice A.

- **Formação acadêmica:** graduações, mestrado, doutorado;
- **Atuações profissionais:** vínculos, atividades de administração, descrição das linhas de pesquisa, atividades de extensão universitária, participações em projetos de pesquisa/extensão/desenvolvimento/outras;
- **Produções bibliográficas:** trabalhos em eventos, artigos publicados, livros publicados ou organizados, capítulos de livros publicados, texto em jornais ou revistas,

¹ Neste exemplo foi utilizado o currículo disponível em <http://lattes.cnpq.br/0378897709136226>

artigos aceitos para publicação, outras produções;

- **Produções técnicas:** patentes, produtos tecnológicos, processos ou técnicas, trabalhos técnicos, apresentação de trabalhos, desenvolvimento de material didático ou instrucional, relatório de pesquisa, outras produções;
- **Orientações concluídas ou em andamento:** mestrado, doutorado, pós-doutorado, iniciação científica, outras;
- **Participação em eventos:** congressos, seminários, simpósios, encontros;

Vários atributos e elementos não foram considerados (dentro do apêndice A existem vários atributos e elementos que estão comentados). Eles parecem conter informações relevantes para determinação da expertise, mas não foram analisados a fundo neste trabalho (em testes preliminares, estes elementos e atributos geraram ruído e, portanto, foram desconsiderados), mas isto não impede de serem avaliados e tratados em trabalhos futuros sob diferentes perspectivas.

4.2.2 Tradução

Como os dados do Currículo Lattes podem estar em idiomas diferentes do Português, é necessário fazer a tradução dos registros para um único idioma, com objetivo de normalizar os dados (Passo 3). A presença de diferentes idiomas ocorre, normalmente, pelo fato das produções estarem representadas no idioma do veículo no qual foram publicadas e, em menor parte, por ter o cadastro de pesquisadores de outros países. A detecção do idioma de origem é feita automaticamente pelo serviço de tradução, discutido na seção .

Como a plataforma Lattes possui quase sua totalidade de currículos referentes a pesquisadores brasileiros, a maior parte das informações estão registradas em Português. Por este motivo, optou-se pela tradução de todas as informações para o Português, com o intuito de diminuir a perda de informação e minimizar o esforço. Deixá-las em seu idioma original poderia trazer 2 problemas: perda de informação relevante e geração de ruído.

4.2.3 Limpeza e normalização

Mesmo se tratando de uma fonte de dados "formal", onde os textos utilizados passaram por revisão gramatical (provavelmente foram revisados durante sua produção e depois por pares ou bancas antes de serem aprovados em algum evento), é necessário realizar diferentes processos de limpeza e tratamento dos textos, permitindo uma recuperação de informação mais eficaz.

A seguir são discutidos os detalhes das cada uma das etapas internas do Passo 4.

4.2.3.1 Limpeza Básica (Passo 4.a)

As entidades HTML são substituídas pelos seus caracteres visuais, iniciando com a sequência `&`, que é transformada em `&`, e depois as entidades no formato `&entidade_nome` ou `&#entidade_numero` pelo representante visual correspondente (por exemplo, a entidade `<` ou sua equivalente `<` é substituída por `<`). Além disso, é feita a remoção de todas as pontuações e outros símbolos. Para este trabalho os seguintes caracteres foram substituídos por espaços: `! " # $ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` { | } ~`.

4.2.3.2 Normalização de palavras (Passo 4.b)

Para este trabalho foi realizada apenas a conversão das letras para minúsculas, assim as comparações e alterações são facilitadas (como é o caso da substituição das *collocations*, discutas adiante, ou agrupamento de palavras no momento da geração dos tópicos).

4.2.3.3 Tratamento dos bigramas (Passo 4.c)

Fazer a identificação de palavras compostas, ou *collocations*, ou ainda bigramas, que são aquelas que ocorrem juntas com frequência (como é o caso de *banco dados*, *raio x*, *sistema operacional*). São palavras que fazem mais sentido quando estão juntas. Para este trabalho foi definido que os bigramas serão calculados para todos os documentos de uma pessoa (unificados), sempre analisando palavras aos pares (também poderiam ser analisadas em trios ou grupos maiores). Serão utilizados aqueles pares que tiverem ocorrido um número mínimo de x vezes. O valor mínimo x é calculado da seguinte forma: $x = m + 2 * DP$, onde m é a média aritmética simples da quantidade de ocorrências dos bigramas e DP é o desvio padrão desse conjunto de valores. Após esse filtro inicial, serão colocadas na ordem decrescente de ocorrência, para que todas sejam substituídas pela palavra unificada com hífen, nessa ordem (passo 4.c). Cada palavra no texto poderá participar de apenas uma substituição. Por exemplo, caso a sequência *ciência política* tenha ocorrido 30 vezes em todos os documentos de uma pessoa e *política internacional* tenha ocorrido 25 vezes, deve-se substituir todas as ocorrências da primeira por *ciência-política* e da segunda por *política-internacional*. Caso encontre em algum documento a sequência *ciência política internacional*, levando em consideração a regra que permite a participação de cada palavra em apenas uma substituição, o trecho ficará *ciência-política internacional*, e não *ciência-política-internacional*. Um passo adicional após substituir as palavras compostas é remover aquelas que ficaram com tamanho inferior a 3, pois são, em sua maioria, sem significância. Palavras que contenham número também são removidas, como datas, URLs, nomes químicos (3-dimetil...), entre outras;

4.2.3.4 Limpeza das palavras desnecessárias globais (Passo 4.d)

Remoção de palavras desnecessárias (ou palavras de ligação), também conhecidas na literatura por *stopwords*. Estas são palavras consideradas não relevantes na análise de textos, justamente por não traduzirem sua essência e normalmente fazem parte desta lista as preposições, pronomes, artigos, advérbios, e outras classes de palavras auxiliares (MORAIS; AMBRÓSIO, 2007). Outras palavras também podem ser adicionadas à lista, como aquelas que aparecem com muita frequência, porém não representam um significado relevante. Considerando o escopo do presente trabalho, a palavra *ufsc* é recorrente nos currículos de pesquisadores vinculados a esta instituição e pode ser adicionada a lista de *stopwords*. Para este trabalho, algumas listas de *stopwords* foram unificadas, totalizando 621 palavras (descritas no Apêndice C).

4.2.3.5 Limpeza individual por currículo (Passo 4.e)

Após a limpeza das palavras desnecessárias globais, que são fixas, é realizada uma estatística da quantidade de palavras para cada currículo, onde o usuário pode selecionar entre as mais recorrentes, quais serão removidas dos seus respectivos currículos. Essa etapa adicional se faz necessária pois podem aparecer palavras específicas do domínio de atuação da pessoa que não possuem grande significância. Nos casos de professores, alguns exemplos são *projeto*, *ensino*, *pesquisa*, *extensão*, entre outras. Essas palavras não fazem parte daquelas removidas no Passo 4.d anteriores e podem fazer sentido em alguns currículos, em outros não, por esse motivo a limpeza é individualizada.

4.2.3.6 Limpeza após agrupamento (Passo 4.f)

Quando vários currículos são analisados em conjunto, algumas palavras que não possuem força contexto no individual, acabam ganhando força maior no conjunto, mas ainda assim não possuem significado relevante. Neste caso, é feita uma nova rodada de contagem de palavras e eliminação das mais recorrentes (escolhidas pelo usuário), mas agora considerando todos os documentos. Por exemplo, a palavra *coordenador* pode aparecer poucas vezes nos currículos individuais e não serão eliminadas no Passo 4.e, mas ao analisar 50 currículos de professores de uma instituição, ela ganha força, mas não confere significados interessantes para a análise da expertise, portanto deve ser eliminada.

4.2.3.7 Outras considerações

Deve-se notar que a aplicação de (4.2.3.2) pode fazer com que palavras percam parte do significado (substantivos próprios viram comuns, siglas e acrônimos ficam mais difíceis de serem reconhecidos) e com (4.2.3.1) palavras compostas (ou siglas pontuadas, como C.E.S.A.R., P&D) são separadas e só serão unificadas na etapa (4.2.3.3) caso tenham ocorrências relevantes. São situações onde a semântica da palavra é alterada (ou até

perdida), o que pode gerar resultados com ruídos. Como não é o objetivo deste trabalho fazer uma análise sobre as técnicas de mineração de textos, esse ponto fica como sugestão para trabalhos futuros.

Um trabalho mais profundo poderia ser realizado no futuro, como pesquisa de gírias, verificação gramatical, correção ortográfica, mesmo que esses problemas ocorram com menor frequência, adicionando ainda a melhoria nos passos (4.2.3.2) e (4.2.3.1).

4.3 Implementação

A implementação do framework foi toda realizada com a linguagem Python, na versão 3.6. A leitura do XML e os processos de limpeza e substituições de ocorrências foram feitos com bibliotecas, expressões regulares ou métodos da própria linguagem, como o pacote `<lxml.etree>` (parser de XML), pacote `<re>` (expressões regulares), pacote `<html>` (tratamento de expressões HTML) e método `<String.replace>` (substituição de ocorrências de caracteres). A localização de bigramas foi implementada utilizando os pacotes `<nltk.collocations.BigramAssocMeasures>` e `<nltk.BigramCollocationFinder>`.

O processo de tradução é feito com o serviço Google Translate, através do uso do pacote `<google.cloud.translate>`, desenvolvido e disponibilizado pela própria empresa. Para ter acesso ao serviço, além de uma conta Google, dentro da sua plataforma² de gerenciamento é necessário criar um projeto, habilitar o serviço *Cloud Translation API* e gerar uma chave privada de acesso, que é um *token* de autenticação (em formato de arquivo). O caminho para o arquivo referente ao *token* deve ser armazenado na variável de ambiente `<GOOGLE_APPLICATION_CREDENTIALS>` antes da execução da aplicação que utiliza o framework.

A empresa desenvolvedora do serviço de tradução limita seu uso por dia, baseando-se no número de caracteres enviados e de requisições feitas. Para os fins deste trabalho, os valores iniciais eram insuficientes (pois a quantidade de traduções é elevada). Neste caso, foram alterados³ para o máximo permitido ainda dentro da faixa gratuita, conforme a lista abaixo. Caso seja necessário traduzir mais caracteres do que o estabelecido abaixo, o usuário deverá optar pelo serviço pago de tradução.

- Characters per day = 1.000.000.000
- Characters per 100 seconds = 10.000.000
- Characters per 100 seconds per user = 10.000.000
- Requests for the list of supported languages per 100 seconds = 1.000

² <https://cloud.google.com/translate/docs/quickstart>

³ Alterados em <https://console.developers.google.com/>, na opção `<API-Quota>`

O idioma de destino escolhido foi o Português para minimizar as perdas decorrentes de traduções, uma vez que a maior parte dos documentos estão nesse idioma (inclusive a maior parte dos currículos pertence a brasileiros, conforme Tabela 2), isto é, menos documentos serão traduzidos (uma vez que a maioria está em Português).

O modelo de referência selecionado para este trabalho, que faz a recuperação da expertise considerando o tempo, foi o Topics Over Time (TOT). Após a transformação dos documentos, tradução e processamento do texto, é este modelo que receberá os documentos compilados para análise da expertise (Passo 5). Sua implementação foi realizada através de duas referências, uma do LDA em Java (NGUYEN, 2018) e outra, *opensource*, em Python, do próprio TOT (MAURYA, 2015). Os gráficos, que auxiliam na interpretação dos resultados do TOT, foram gerados com a biblioteca `<matplotlib>` (Passo 6).

Foram implementadas diversas classes para contemplar os 6 passos do framework, descritos na Figura 13, que podem ser usadas de maneira separada pelo usuário, ou através da classe principal `TemporalAnalysis`, que faz o correto encadeamento do fluxo de execução. Esta classe recebe como parâmetro as seguintes informações:

- Uma lista com os caminhos dos arquivos XML a serem analisados;
- Uma flag booleana indicando se os dados devem ser analisados do início ou se podem ser utilizados os dados do cache;
- O caminho do arquivo JSON contendo a estrutura do currículo Lattes a ser considerada;
- Os parâmetros específicos do TOT:
 - Valor de α ;
 - Valor de β ;
 - Quantidade de iterações do amostrador de Gibbs;
 - Quantidade de tópicos;
- Lista com os caminhos dos arquivos texto contendo as palavras desnecessárias globais;
- O número de palavras para serem exibidas ao usuário durante o filtro individual por currículo;
- O caminho para o arquivo JSON contendo as credenciais Google para acessar o serviço de tradução;
- O ano inicial e final para filtrar os documentos que serão submetidos ao TOT;
- O caminho para o arquivo onde os dados do resultado serão gravados;
- Um objeto responsável pelo log do sistema.

Com todas essas informações, o framework é capaz de realizar os passos previstos e apresentar ao usuário a análise da expertise de um ou vários currículos ao mesmo tempo.

5 Experimentos e resultados

Nesse capítulo são realizados os experimentos necessários para teste e validação do framework, com definição dos objetivos, metodologia usada, métricas de avaliação e, por fim, análise de resultados.

5.1 Objetivos

Para validação da corretude do framework proposto, foram executados dois tipos de experimentos: análise individual de currículos, mostrando dados mais específicos do pesquisador, e análise de um grupo de currículos, mostrando dados macro (e mais abrangentes) do grupo. Dessa maneira, os objetivos da etapa de testes podem ser divididos da seguinte maneira:

- Analisar currículos de maneira individual, com o objetivo de validar a identificação da expertise de um pesquisador e como ela evolui ao longo do tempo, confrontando os resultados do framework com os dados cadastrados na Plataforma Lattes; e
- Analisar grupos de pesquisadores, com o objetivo de validar o comportamento do framework com a presença de uma quantidade maior de dados, tanto no reconhecimento dos tópicos e sua evolução ao longo do tempo, bem como o tempo de execução.

5.2 Metodologia

Para cumprir com os objetivos elencados, é definida uma metodologia para execução dos experimentos, que envolve a seleção de Programas de Pós Graduação, escolha de pesquisadores para análise individual e definição dos parâmetros de execução.

Nesse sentido, a Plataforma Lattes foi acessada e foram baixados (via script) todos os currículos (em formato XML) dos pesquisadores credenciados nos Programas de (i) Pós Graduação em Ciência da Computação (PPGCC), (ii) Administração (PPGADM/CPGA) e (iii) Administração Universitária (PPGAU). Estes programas foram escolhidos a partir da familiaridade do autor com os pesquisadores credenciados em cada programa (o que facilita a análise posterior dos dados). Os currículos do PPGCC foram colocados em um grupo; PPGAU e PPGADM foram unificados (eliminando os duplicados) e colocados em um segundo grupo. A unificação foi realizada pois se tratam de áreas correlatas (pertencentes a mesma subárea do conhecimento), possuindo, inclusive, alguns membros em comum no corpo docente. A escolha dos currículos e parametrização para execução dos testes são descritas a seguir.

5.2.1 Experimentos com currículos individuais

Para validar o framework com os currículos de maneira individual, foram selecionados 3 currículos nos dois grupos de pesquisadores formados acima (um grupo do PPGCC e outro grupo formado por PPGAU e PPGADM). A única restrição foi escolher currículos entre os maiores (considerando o tamanho do arquivo XML, dado que há interesse no volume de dados), pois foi adotada a premissa de quanto maior o arquivo, mais informações estão cadastradas.

Nesta etapa, foram selecionados 2 currículos de pesquisadores do PPGCC, da área de Ciência da Computação e 1 do PPGADM, que é da área de Economia (mas está credenciado no programa da Administração). Os IDs dos currículos e sua referência deste ponto em diante podem ser conferidos na Tabela 5.

| Referência | ID Lattes | Área |
|------------|------------------|------------|
| <i>a</i> | 0378897709136226 | Computação |
| <i>b</i> | 5509228706749288 | Economia |
| <i>c</i> | 4386203755404108 | Computação |

Tabela 5 – IDs dos currículos utilizados nos testes

Fonte: Elaborado pelo autor.

Todos os currículos foram submetidos individualmente ao TOT, utilizando a parametrização inicial definida na Tabela 6. Além disso, os tópicos de cada palavra, de cada documento, são inicializados aleatoriamente entre os T disponíveis. Vale ressaltar que os parâmetros ψ_z , θ_d e ϕ_z são justamente aqueles ajustados em cada iteração do Amostrador de Gibbs (os demais permanecem constantes). Alguns autores discutem formas de calcular o valor ideal para o número de tópicos, mas para este trabalho optou-se pelo valor fixo de 6. No trabalho de (WANG; MCCALLUM, 2006), onde o TOT é definido, o número de tópicos é definido como 50, também como um valor fixo, pois o conjunto de dados analisado era extenso (cartas trocadas entre o Presidente dos Estados Unidos e o Congresso americano, entre os anos de 1790 e 2002). Não há uma discussão aprofundada sobre esse número *mágico*, mas acredita-se ser um valor para mera análise do conjunto de dados, com objetivo de descobrir quais os tópicos latentes que poderiam surgir. No caso deste trabalho, o valor 6 foi considerado para tentar capturar possíveis mudanças na expertise do pesquisador, levando em conta poucas alterações em seu campo de pesquisa. Foram realizados testes com outros números, mas quando este é muito elevado, os tópicos perdem o sentido (misturando palavras com pouco significado), deixando palavras importantes esparsas; quando é muito baixo, os tópicos mais *fortes* (aqueles com palavras que ocorrem muito) tomam conta e praticamente todos ficam com o mesmo conteúdo. A quantidade 6 pareceu a mais razoável para os dados analisados.

| Parâmetro | Valor inicial |
|----------------------------------|-------------------|
| α | 0.1 |
| β | 0.1 |
| ψ_z (tópico x tempo) | $(a, b) = (1, 1)$ |
| θ_d (documento x tópico) | 0 |
| ϕ_z (palavra x tópico) | 0 |
| Iterações do amostrador de Gibbs | 1000 |
| T (tópicos por currículo) | 6 |
| Ano inicial | 0 |
| Ano final | 2018 |

Tabela 6 – Parâmetros utilizados na inicialização do TOT.

Fonte: Elaborado pelo autor.

5.2.2 Experimento com grupo de pesquisadores

Para analisar um grupo de pesquisadores, foi escolhido o PPGCC, também pela familiaridade do autor com os membros do programa. Neste sentido, o trabalho considerou todos aqueles que constavam no Sistema de Controle Acadêmico da Pós-Graduação (CAPG) como membros do programa, até agosto de 2018, sendo coletados 27 currículos.

Todos foram individualmente processados pelos passos iniciais (geração, limpeza e tratamento de bigramas), sendo que todos os documentos de todos os currículos foram agrupados para gerar uma nova etapa de limpeza (referente ao Passo 4.f do framework, referente a limpeza após agrupamento). Esse resultado foi enviado para execução do TOT (representado pelo Passo 5), para prosseguimento da execução. Como a característica do conjunto dados é diferente da análise individual (com mais dados e mais campos de pesquisa envolvidos), os parâmetros de execução foram alterados, conforme pode ser verificado na Tabela 7. O valor de α mais alto se deve ao fato de que, sabidamente, existem mais tópicos no conjunto de currículos analisados (comparado a uma análise individual), portanto, os documentos provavelmente possuem uma mistura maior dos tópicos existentes. Como existem mais palavras no conjunto analisado, considerou-se um valor menor de β , para que os tópicos tenham uma mistura menor das palavras disponíveis. O número de tópicos foi considerado como $T = 10$, uma vez que existem 7 linhas de pesquisa dentro do programa, o valor foi levemente aumentado para comportar possíveis variações dentro do algoritmo. Valores menores que 7 não fazem sentido, escondendo tópicos ou forçando uma relação irreal na composição. Valores muito altos sofrem do mesmo problema relatado no caso das análises individuais (palavras ficam esparsas e tópicos sem sentido são criados). Valor igual a sete também não permite que tópicos um pouco mais fracos apareçam (por exemplo: dois tópicos fortes de Banco de Dados podem tomar o lugar de um tópico de Computação Paralela e Distribuída). Durante os experimentos, o valor 10 se mostrou mais adequado para o conjunto de dados.

O período de início e fim foi considerado o mesmo da criação do programa até o momento do início desta pesquisa. Mesmo que os atuais integrantes provavelmente não sejam os mesmos de 1992, é importante verificar se eles estiveram e continuam aderentes às linhas de pesquisa. Todos os valores foram aproximados a partir de testes realizados com os dados, que foram considerados aceitáveis para o conjunto de dados.

| Parâmetro | Valor inicial |
|----------------------------------|-------------------|
| α | 0.15 |
| β | 0.03 |
| ψ_z (tópico x tempo) | $(a, b) = (1, 1)$ |
| θ_d (documento x tópico) | 0 |
| ϕ_z (palavra x tópico) | 0 |
| Iterações do amostrador de Gibbs | 1000 |
| T (quantidade de tópicos) | 10 |
| Ano inicial | 1992 |
| Ano final | 2018 |

Tabela 7 – Parâmetros utilizados na inicialização do TOT na execução para os currículos do PPGCC.

Fonte: Elaborado pelo autor.

5.3 Métricas de avaliação

Com o objetivo de analisar os resultados, são definidas 3 ferramentas importantes que exibem os dados da execução do framework: *Mapa de calor* (mostra as principais palavras de cada tópico), a *Evolução dos tópicos* (mostra a evolução dos tópicos ao longo do tempo) e a *Probabilidade posterior acumulada* (uma aproximação para complementar a evolução dos tópicos). Cada uma é discutida adiante.

5.3.1 Mapa de calor

A execução do TOT em cada um dos currículos definidos na Tabela 5, ou para o grupo de pesquisadores, estima os valores dos parâmetros ϕ , θ e ψ para cada um dos T tópicos. As probabilidades definidas por ϕ indicam qual a probabilidade da palavra w ocorrer quando se trata do tópico $t \in T$. A soma de todas as probabilidades das palavras de um tópico é igual a 1. Como os tópicos definidos pelo TOT não são mutáveis (diferentemente do TET), foram gerados mapas de calor indicando as 5 palavras de maior probabilidade de cada tópico nos casos individuais e as 12 palavras no caso do grupo. Esse valor indica quanto o tópico *gosta* da palavra em questão e aqui pode ser conferida a *composição semântica* de cada um. Quando uma palavra aparece no topo, isso indica que ela tem a maior probabilidade para ocorrer em documentos desse tópico. A numeração

sequencial dos tópicos é meramente ilustrativa e seu nome deve ser atribuído por quem analisa os dados.

5.3.2 Evolução dos tópicos

Com relação às ocorrências de cada tópico ao longo do tempo, definidas pela variável ψ (que é uma distribuição Beta), foram gerados os gráficos 15, 18, 21 e 24, que apresentam sua evolução. Esta informação indica em quais períodos houve maior incidência de cada tópico, bem como o surgimento ou decaimento de alguma nova linha de pesquisa e a provável expertise do pesquisador em cada período. Essa análise precisa ser atrelada com a composição semântica do tópico: o TOT é mais sensível do que o LDA, portanto uma pequena mudança na coocorrência das palavras pode gerar um novo tópico, isto é, um tópico envolvendo *reconhecimento de padrões em imagens médicas*, que deixou de ter incidência nos últimos anos e outro *reconhecimento de padrões em imagens por satélite*, que vem emergindo, ainda pode indicar uma expertise na área de *visão computacional*.

5.3.3 Probabilidade posterior acumulada

Como cada documento é considerado como uma mistura de vários tópicos e como essa informação é apurada em θ (a probabilidade de cada tópico compor o documento em questão), é possível corroborar os gráficos de evolução com os de probabilidade acumulada, de maneira aproximada, com a separação das probabilidades pelo ano do documento. Isto é, para cada tópico, é feito o somatório de todas as probabilidades vinculadas a ele, de todos os documentos de um determinado ano. A soma da probabilidade de todos os tópicos possíveis em um documento é sempre igual a 1, porém, no gráfico de probabilidades acumuladas, como é feito o somatório de todos os valores por tópico X ano X documento, podem existir valores maiores que 1. De outra maneira: os documentos são formados por uma mistura de tópicos, havendo uma probabilidade de que o documento pertença a cada um deles (essa é a informação apurada em θ). Se em uma situação existem 3 tópicos t_1 , t_2 e t_3 e o TOT apura para um documento d as probabilidades $\theta_{d1} = 0.6$, $\theta_{d2} = 0.3$ e $\theta_{d3} = 0.1$, isso indica que a probabilidade do documento d ser formado por palavras do tópico 1 é de 60%, do tópico 2 de 30% e do tópico 3 de 10%. O somatório dessas probabilidades para todos os documentos agrupados por ano são exibidos nos gráficos de probabilidades posteriores acumuladas. Apesar dessa medida não ser prática, os valores estão normalizados e espera-se que os gráficos de evolução e de probabilidades acumuladas sejam próximos.

5.4 Análise dos resultados

Os resultados obtidos através da execução do framework para os conjunto de dados definidos nesse capítulo são discutidos nas seções seguintes.

5.4.1 Experimentos com currículos individuais

5.4.1.1 Currículo *a*, da área de Ciência da Computação

Para o currículo *a*, da área de Ciência da Computação, foram gerados $D = 246$ documentos, sendo que a etapa de tradução detectou 4 idiomas (Português, Inglês, Galego e Espanhol). Os idiomas galego e espanhol provavelmente foram erros da API, porém não alteraram significativamente o resultado. Os dados, com amostras da tradução, podem ser conferidos na Tabela 8. Na etapa de limpeza individual do currículo, foram removidas as seguintes palavras: *abordagem, acesso, análise, aplicações, áreas, baseado, biblioteca-digital, bibliotecas-digitais, característica, características, científica, colaboração, comunidade, contexto, controle, coordenação, criação, desenvolvimento, geração, método, modelagem, modelo, objetivo, objeto, páginas, perfil, pesquisadores, principal, problemas, processo, produção-científica, projeto, sistemas, sul, suporte, tabelas, técnica, técnicas, trigo, uso, usuário, usuários e volume.*

| Idioma | Qt docs | Amostras da tradução | |
|-----------|---------|---|---|
| Português | 214 | | |
| Inglês | 27 | Research on Complex Data Management and Analysis at UFSC. Journal of Information and Data Management - JIDM | Pesquisa em Gestão e Análise de Dados Complexos na UFSC. Jornal de Informação e Gestão de Dados - JIDM |
| | | Handling temporal information in web search engines. SIGMOD Record3 | Manipulando informações temporais em mecanismos de pesquisa da web. Registro SIGMOD |
| | | Improving Performance in a DBaaS Environment through the use of Resource Reservation5 | Melhorando o desempenho em um ambiente DBaaS por meio do uso de reserva de recursos |
| Galego | 3 | Interoperabilidade de sistemas e dados utilizando SOA e BPEL. Interoperabilidade de sistemas. SOA. BPEL. Web Services | Interoperabilidade de sistemas e dados usando SOA e BPEL. Interoperabilidade de sistemas. SOA BPEL. Serviços da Web |
| | | Interoperabilidade entre microformatos | Interoperabilidade entre microformatos |
| | | Aplicando os Recursos de XML | Aplicando os Recursos XML |
| Espanhol | 2 | Clustering de documentos XML. XML. Algoritmos de clustering | Cluster de documentos XML. XML Algoritmos de Cluster |
| | | Busca de microformatos em páginas Web | Pesquisa de microformatos em páginas da Web |

Tabela 8 – Idiomas detectados no currículo *a*, com amostras de traduções realizadas
Fonte: Elaborado pelo autor.

Com relação aos bigramas, foram encontrados 2242 possíveis palavras compostas, sendo que apenas as 52 com maior ocorrência foram substituídas. As estatísticas usadas para o corte estão na Tabela 9, e as 5 maiores podem ser conferidas na Tabela 10.

| Medida | Valor |
|--|------------|
| Quantidade de bigramas distintos | 2247 |
| Total de ocorrências | 2916 |
| Média | 1,299 |
| Variância | 1,644 |
| Desvio padrão | 1,282 |
| Média + 2 * Desvio padrão | 3,864 |
| Bigramas distintos com mais de 3,864 ocorrências | 52 (2,31%) |

Tabela 9 – Tabela com estatísticas sobre os bigramas do currículo *a*.

Fonte: Elaborado pelo autor.

| Bigramas | Substituído por | Ocorrências substituídas |
|-------------------|-------------------|--------------------------|
| banco dados | banco-dados | 38 |
| integração dados | integração-dados | 17 |
| semi estruturados | semi-estruturados | 16 |
| dados semi | dados-semi | 15 |
| extração dados | extração-dados | 14 |

Tabela 10 – Exemplos de bigramas substituídos nos documentos do currículo *a*.

Fonte: Elaborado pelo autor.

Para o currículo *a*, é possível verificar na Figura 14 os 6 tópicos e as principais palavras de sua composição. Nesse e nos demais mapas de calor, as palavras não estão distribuídas no tempo (mas o tempo influenciou sua localização dentro dos tópicos). A palavra *dados* é uma das que ocorre com muita frequência, coocorrendo com: *web* e *ambiente*, *representações* (tópico 1); *xml*, *similaridade*, *integração* (tópico 2); *web*, *conhecimento*, *gerenciamento* (tópico 3); *integração*, *web-services*, *casamento-aproximado* (tópico 5); e *web*, *extração*, *documentos* (tópico 6). Percebe-se aqui a correlação entre termos: *xml* é sabidamente um formato amplamente utilizado na *web*, usado inclusive em serviços SOAP ou REST, consumidos por *web-services*, para *representação* de dados de maneira semi-estruturada, sendo que existem *bancos de dados* especialistas nesse formato. Essas informações indicam expertise em diferentes frentes relacionadas a *gerenciamento de dados*. Essa análise casa com trecho informado pelo próprio autor, em seu currículo: *Seus interesses de pesquisa incluem as áreas de Gerenciamento de Dados, Recuperação de Informação, Mineração de Dados com ênfase na Web, Descoberta de Conhecimento e Extração e Matching de Informação*. Com relação à análise no tempo, os tópicos 1, 4 e 6 são os mais proeminentes nos últimos anos, conforme verificado no gráfico da Figura 15, que indicam uma tendência para trabalho com extração e similaridade de dados, provavelmente ligados à *web*. Alguns documentos envolvendo esses dois tópicos, com seu ano, podem ser conferidos na Tabela 11.

| Ano | Documento (parcial) |
|------|---|
| 2000 | Extração de Dados XML com base em uma Ontologia |
| 2008 | Um Modelo Temporal para Definição do Perfil de Pesquisadores Utilizando Dados do Lattes |
| 2012 | IPU: Coleta, a organização e publicação dos dados da produção científica e técnica dos docentes da UFSC [...] Este projeto tem como principal objetivo o mapeamento de grandes volumes de dados XML para persistência em um banco de dados relacional. O foco é realizar a alimentação do banco de dados com dados públicos contidas nos currículos Lattes dos membros da comunidade universitária, e uma plataforma web de apoio à tomada de decisões por parte dos gestores para assuntos como a produção bibliográfica, técnica e artística dos docentes da instituição. Os dados são coletados através de um webservice fornecido pelo CNPq, tratados através de uma ferramenta de parse dos arquivos XML recebidos e padronizados para a inserção no banco de dados elaborado no projeto |
| 2012 | Uma Abordagem para Detecção e Extração de Rótulos em Formulários Web |
| 2015 | Sistemas de Integração Lattes-Qualis |
| 2016 | Agrupamento aproximado de Currículos Lattes por Afinidade de Áreas de Conhecimento Considerando Temporalidade |
| 2016 | Agrupando aproximado de pesquisadores por co-autoria de publicações segundo o currículo Lattes |
| 2017 | Mariage: Extração e casamento de dados estruturados da Web com múltiplas representações |
| 2017 | Extração de Informações de Conferências em Páginas Web |

Tabela 11 – Exemplos de produções extraídas diretamente da página do currículo *a*
Fonte: Adaptado do Currículo Lattes

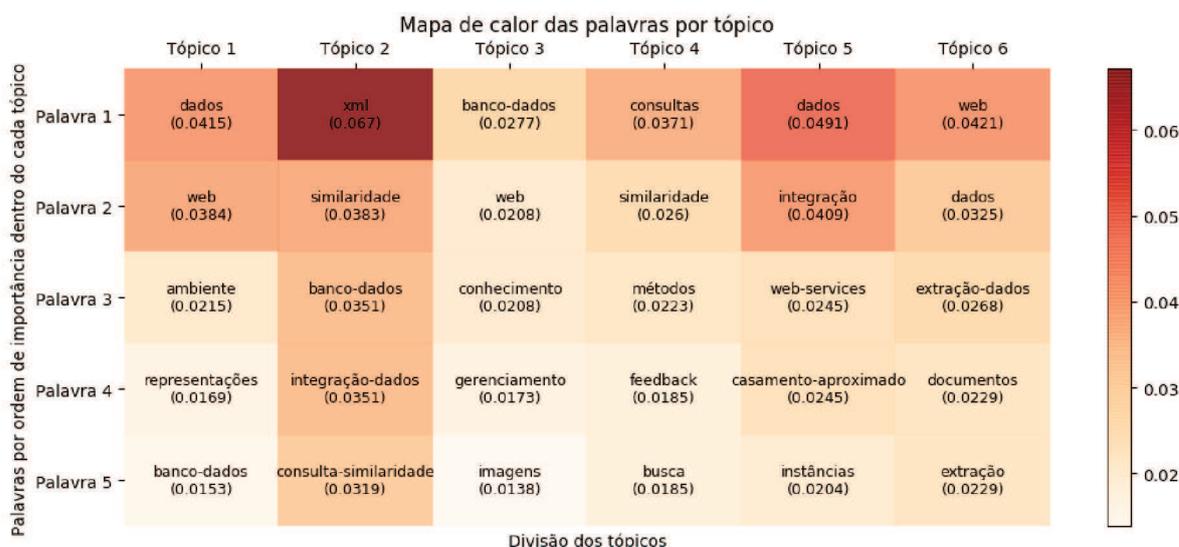


Figura 14 – Mapa de calor das palavras por tópico do currículo *a*
Fonte: Elaborado pelo autor.

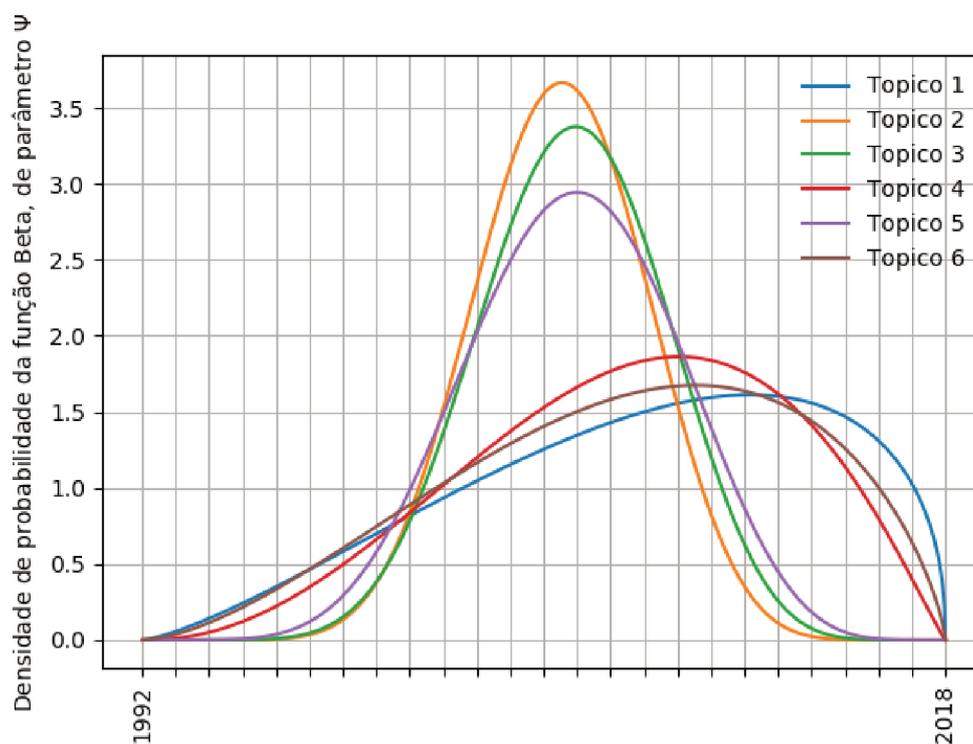


Figura 15 – Evolução dos tópicos ao longo do tempo do currículo *a*

Fonte: Elaborado pelo autor.

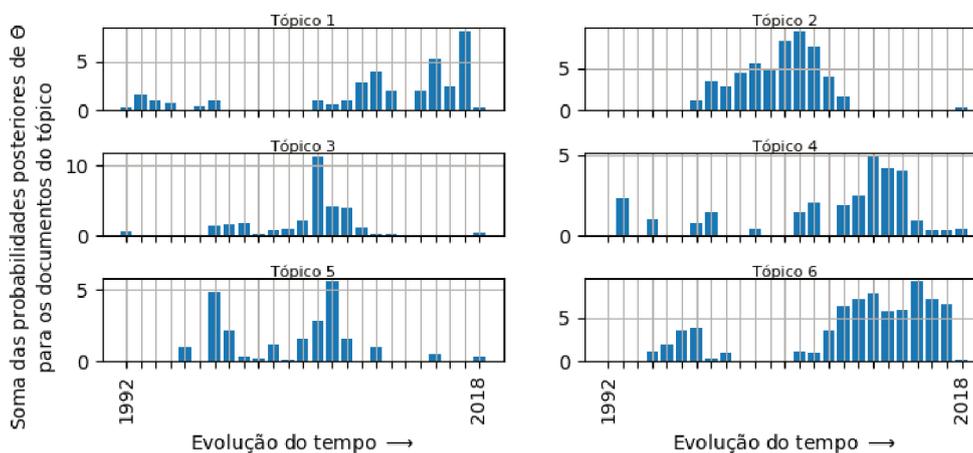


Figura 16 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo *a*

Fonte: Elaborado pelo autor.

5.4.1.2 Currículo *b*, da área de Economia

Já para o currículo *b*, da área de Economia, foram gerados $D = 761$ documentos, sendo que a etapa de tradução detectou 5 idiomas (Português, Inglês, Espanhol, Galego e

Romeno). Os idiomas galego e romeno foram considerados como erros de reconhecimento API de tradução. O idioma espanhol, entretanto, apresentou algumas traduções válidas e outras praticamente sem alteração (muito provável pela proximidade do português). Os dados, com amostras da tradução, podem ser conferidos na Tabela 12. Na etapa de análise individual de palavras desnecessárias, foram filtrados os seguintes termos: *ambiente, análise, branco, brasileira, caso, década, economia, evidências evolução, período, pluriatividade, política, políticas, pós, produção, programa, projeto, revista, santa-catarina, século-xxi, setor, social, texto-discussão e ufsc*.

| Idioma | Qt docs | Amostras da tradução | |
|-----------|---------|--|--|
| Português | 731 | | |
| Inglês | 17 | New Socioeconomic Dimensions of the Brazilian rural space | Novas dimensões socioeconômicas do espaço rural brasileiro |
| | | The Brazilian rural development model in the context of green economy: myths and reality. Green Economy | O modelo de desenvolvimento rural brasileiro no contexto da economia verde: mitos e realidade. Economia verde |
| | | Agrarian Reform in Brazil under Neoliberalism: Evaluation and Perspectives. Reforma agraria - politicas publicas | Reforma agrária no Brasil sob o neoliberalismo: avaliação e perspectivas. Reforma agraria - politicas publicas |
| Espanhol | 11 | Por una América Latina sin hambre: politicas de seguridad alimentaria. América Latina. Fome | Para uma América Latina sem fome: políticas de segurança alimentar. América Latina. Fome |
| | | Reforma agraria en Brasil bajo el neoliberalismo: evaluación y perspectivas. Questao agraria. Brasil | Reforma agrária no Brasil sob neoliberalismo: avaliação e perspectivas. Questao Agrícola. Brasil |
| | | Turismo rural | Turismo rural |
| | | América Latina. Pobreza. Política pública | América Latina. Pobreza Política pública |
| Galego | 1 | Membro do corpo de pareceristas da Revista da SEP. | Membro do corpo de juizes da Revista SEP. |
| Romeno | 1 | Halloween. Saci | Dia das Bruxas. sacos |

Tabela 12 – Idiomas detectados no currículo *b*, com amostras de traduções realizadas
Fonte: Elaborado pelo autor.

Com relação aos bigramas, foram encontradas 4965 possíveis palavras compostas, sendo que apenas as 48 com maior ocorrência foram substituídas. As estatísticas usadas

para o corte estão na Tabela 13, e as 5 maiores podem ser conferidas na Tabela 14.

| Medida | Valor |
|---|------------|
| Quantidade de bigramas distintos | 4965 |
| Total de ocorrências | 8675 |
| Média | 1,783 |
| Variância | 2,109 |
| Desvio padrão | 4,702 |
| Média + 2 * Desvio padrão | 11,188 |
| Bigramas distintos com mais de 11,188 ocorrências | 48 (0,96%) |

Tabela 13 – Tabela com estatísticas sobre os bigramas do currículo *b*.

Fonte: Elaborado pelo autor.

| Bigramas | Substituído por | Ocorrências substituídas |
|-----------------------|-----------------------|--------------------------|
| santa catarina | santa-catarina | 257 |
| agricultura familiar | agricultura-familiar | 94 |
| desenvolvimento rural | desenvolvimento-rural | 75 |
| américa latina | américa-latina | 53 |
| século xx | século-xx | 53 |

Tabela 14 – Exemplos de bigramas substituídos nos documentos do currículo *b*.

Fonte: Elaborado pelo autor.

No caso do currículo *b*, as palavras que compõem cada um dos tópicos podem ser conferidas no mapa de calor da Figura 17. Percebe-se uma grande relação das palavras com situações econômicas de vulnerabilidade, principalmente vinculadas à: *fome, pobreza, agricultura-familiar, desenvolvimento-rural* e outras áreas rurais. Como este é um currículo da área de Ciências Econômicas, é provável que as relações entre fome, pobreza e emprego são analisados em níveis econômicos, analisando políticas públicas e o desenvolvimentismo e, pelos termos que indicam locais, parece que o estudo é direcionado para Santa Catarina e America Latina. Essas informações casam com a descrição dada pelo próprio autor em sua página do Currículo Lattes: *Atua na área de Economia, priorizando os seguintes temas: desenvolvimento econômico, desenvolvimento rural e territorial, economia brasileira e regional, economia do trabalho e políticas públicas de erradicação da pobreza e da exclusão social. Coordenador do NECAT (Núcleo de Estudos de Economia Catarinense)*. Os tópicos 4 e 5 parecem ser os mais atuais na pesquisa, estando relacionados com a pobreza rural. Através do Gráfico 19, é possível perceber que houve uma queda na produção em praticamente todos os tópicos próximo ao final do período de análise (com exceção do tópico 5), sendo que todos tiveram uma retomada considerável no final. Na Tabela 15 podem ser verificados alguns documentos com seus respectivos anos

| Ano | Documento (parcial) |
|--------------|--|
| 2009 | Efeitos do Programa de Aquisição de Alimentos da Agricultura Familiar (PAA) em municípios do Planalto norte do estado de Santa Catarina. |
| 2011 | Aperfeiçoamento das políticas públicas de enfrentamento da pobreza rural na perspectiva do desenvolvimento territorial |
| 2012 | Pobreza rural: um fenômeno histórico-estrutural relacionado à estrutura agrária do país |
| 2015 | Políticas sociais de enfrentamento da pobreza na América Latina: uma análise comparada entre Brasil e Argentina. |
| 2015 | Análise de impactos do PRONAF em municípios delecionados de Santa Catarina. |
| 2016 2016 | A importância do sistema familiar de produção no estado de Santa Catarina Articulação de políticas para a superação da pobreza rural: um estudo comparativo das interfaces entre programa Bolsa Família e as políticas de inclusão produtiva nas regiões Nordeste e sul do Brasil |
| 2017 | Desigualdade e pobreza na América Latina: trajetória histórica e contexto atual. Descrição: O projeto buscar analisar os cenários da desigualdade e da pobreza na América Latina de forma simultânea e conectada, porém destacando as múltiplas dimensões de cada um desses fenômenos em particular. [...] |
| 2018 | Emprego e Desemprego em tempos de crise econômica e reforma trabalhista. |

Tabela 15 – Exemplos de produções extraídas diretamente da página do currículo *b*
Fonte: Adaptado do Currículo Lattes.

Sobre a temporalidade deste currículo, percebe-se que a maior parte dos documentos ocorrem de 1999 em diante, o que casa com o término do doutorado deste pesquisador (também em 1999). Os tópicos mais proeminentes são o 4 e o 5. Ambos envolvem muito o conteúdo produzido pelo Núcleo de Estudos de Economia Catarinense (NECAT), coordenado pelo pesquisador deste currículo. Realizando uma pesquisa no Diretório dos Grupos de Pesquisa no Brasil, a informação é de que este núcleo foi criado em 2012. Essas informações levam a uma expertise na área socioeconômica, envolvendo assuntos como combate a pobreza, políticas públicas e desenvolvimento rural, direcionados para as regiões da América Latina (macro) e Santa Catarina (micro).

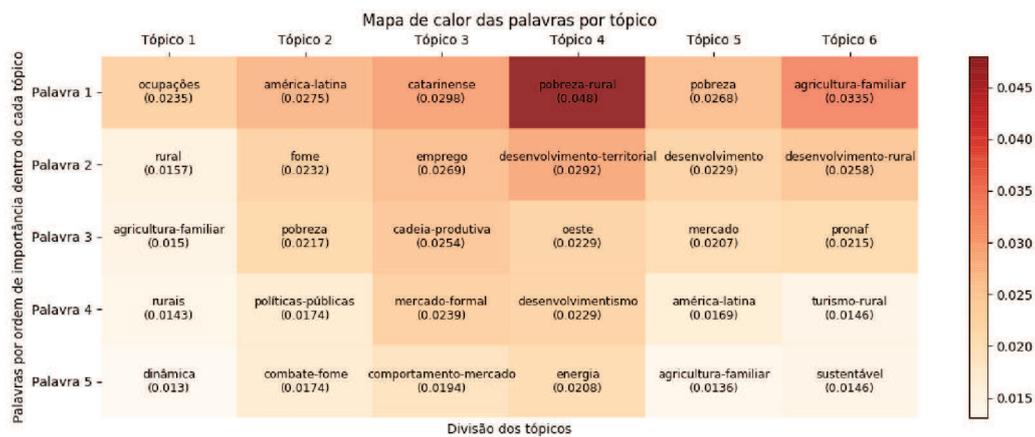


Figura 17 – Mapa de calor das palavras por tópico do currículo *b*
 Fonte: Elaborado pelo autor.

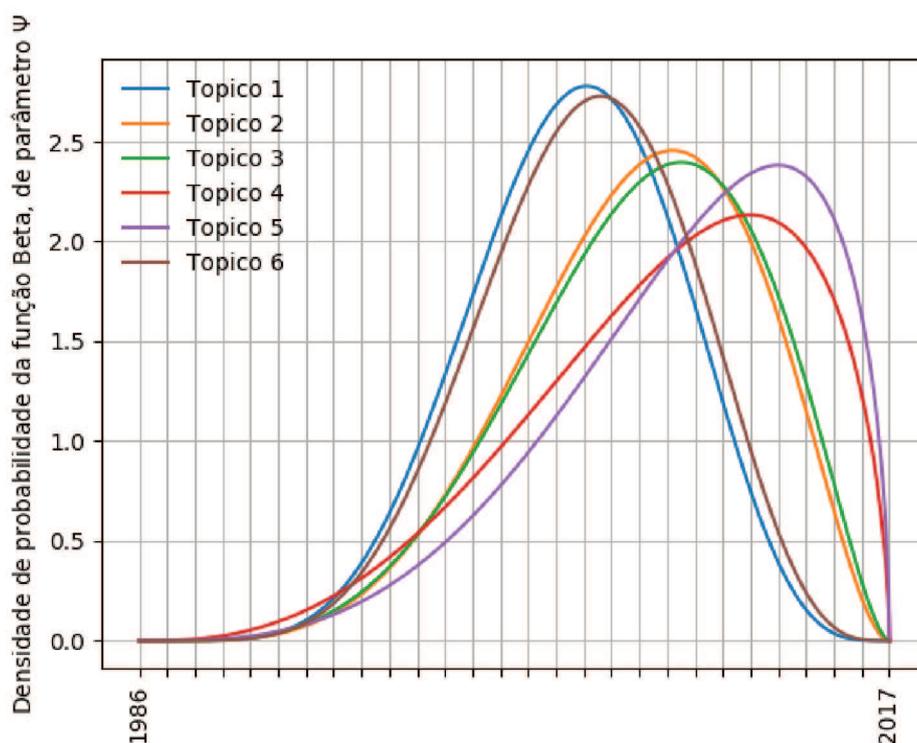


Figura 18 – Evolução dos tópicos ao longo do tempo do currículo *b*
 Fonte: Elaborado pelo autor.

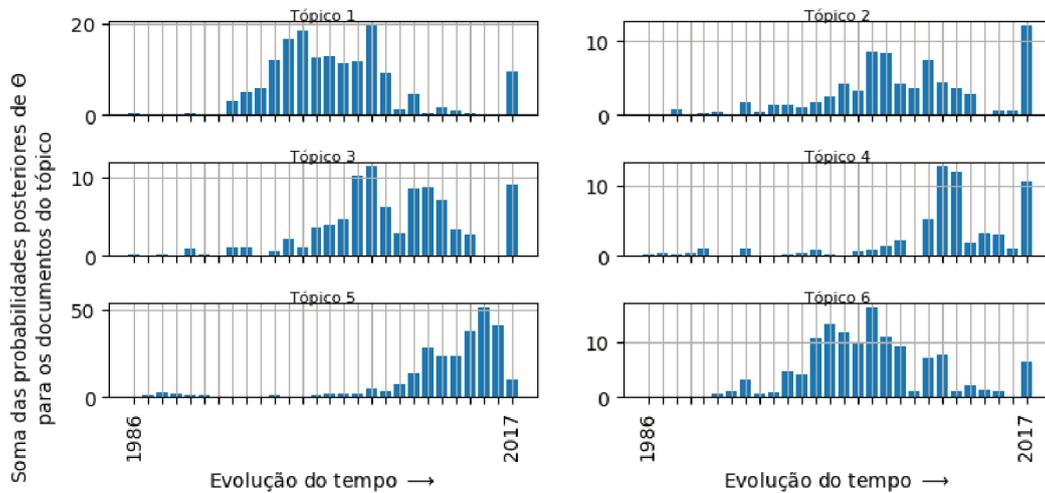


Figura 19 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo b

Fonte: Elaborado pelo autor.

5.4.1.3 Currículo c , da área de Ciência da Computação

No caso do currículo c , da área de Ciência da Computação, foram gerados $D = 736$ documentos, sendo que a etapa de tradução detectou 6 idiomas (Português, Inglês, Francês, Espanhol, Galego e Hmong). Os idiomas galego e hmong foram considerados como erros de reconhecimento API. Os dados, com amostras da tradução, podem ser conferidos nas Tabelas 16 e 17. Na etapa da análise individual do currículo, foram filtradas as palavras *análise*, *and*, *apresentador*, *baseado*, *chair*, *comitê*, *conferência-internacional*, *desenvolvimento*, *ids*, *ieee*, *ieee-cnom*, *management*, *membro*, *modelo*, *projeto*, *proposta*, *rede*, *revista*, *secretário*, *service*, *tutorial*, *uids* e *uso*

| Idioma | Qt docs | Amostras da tradução | |
|-----------|---------|---|--|
| Português | 488 | | |
| Inglês | 235 | Hybrid Protocol for Group Key Management and Transparent Cluster Interconnection in a Secure Wireless Sensor Network | Protocolo híbrido para gerenciamento de chaves de grupo e interconexão de cluster transparente em uma rede de sensores sem fio seguros |
| | | Challenges in Cloud Computing and IoT Security | Desafios na computação em nuvem e na segurança da IoT |
| | | Analytical Model to Evaluate the Performance of Mobile Agents in a Generic Network Topology. Journal of Network and Systems Management | Modelo analítico para avaliar o desempenho de agentes móveis em uma topologia de rede genérica. Jornal de Gestão de Redes e Sistemas |
| Francês | 4 | Conception et développement de l'architecture d'administration d'un réseau métropolitain. gerência de redes. Redes Metropolitanas. Redes de Computadores. Redes de Telecomunicações. TMN. OSI | Projeto e desenvolvimento da arquitetura administrativa de uma rede metropolitana. gerência de redes. Redes Metropolitanas. Redes de Computadores. Redes de Telecomunicações. TMN. OSI |
| | | Biblotos: Une Bibliotheque de Constructions Predefinies Pour Un Projet de Gestion de Reseau | Biblotos: Uma biblioteca de construções pré-definidas para um projeto de gerenciamento de rede |
| | | Connexions donnés, voix et image dans l'administration de réseaux hétérogènes. Les Dossiers Des Réseaux D'entreprise. gerência de redes. Vox. Dados. Imagem. Redes Heterogêneas | Conexões dadas, voz e imagem na administração de redes heterogêneas. Arquivos de Rede Corporativa. gerência de redes. Vox. Dados. Imagem. Redes Heterogêneas |

Tabela 16 – Idiomas detectados no currículo *C*, com amostras de traduções realizadas (*parte 1*)

Fonte: Elaborado pelo autor.

| Idioma | Qt docs | Amostras da tradução | |
|----------|---------|---|---|
| Espanhol | 4 | Proyeto Formal de Una Plataforma de Gerenciamento. Gerencia de Redes | Projeto formal de uma plataforma de gerenciamento. Gerenciamento de Rede |
| | | Gerencia Distribuida Na Area de Seguranca Para Redes de Telecomunicacoes | Gestão Distribuída Na Segurança Área de Redes de Telecomunicações |
| Galego | 3 | Foro plenário sobre educación en ingeniería informática | Fórum Plenário de Educação em Engenharia da Computação |
| | | Realizacao da Gerencia Distribuida de Redes Utilizando Snmp, Java, Www e Corba. Gerencia de Redes | Implementação do Gerenciamento Distribuído de Redes utilizando Snmp, Java, Www e Curva. Gerenciamento de Rede |
| Hmong | 2 | QoS em Wireless | QoS em Wireless |

Tabela 17 – Idiomas detectados no currículo *c*, com amostras de traduções realizadas (*parte 2*)

Fonte: Elaborado pelo autor.

Com relação aos bigramas, foram encontradas 3688 possíveis palavras compostas, sendo que apenas as 43 com maior ocorrência foram substituídas. As estatísticas usadas para o corte estão na Tabela 18, e as 5 maiores podem ser conferidas na Tabela 19.

| Medida | Valor |
|---|------------|
| Quantidade de bigramas distintos | 3688 |
| Total de ocorrências | 6507.0 |
| Média | 1,764 |
| Variância | 20,138 |
| Desvio padrão | 4,487 |
| Média + 2 * Desvio padrão | 10,739 |
| Bigramas distintos com mais de 11,188 ocorrências | 43 (1,16%) |

Tabela 18 – Tabela com estatísticas sobre os bigramas do currículo *c*.

Fonte: Elaborado pelo autor.

| Bigramas | Substituído por | Ocorrências substituídas |
|-------------------|-------------------|--------------------------|
| gerência rede | gerência-rede | 223 |
| computação nuvem | computação-nuvem | 89 |
| rede fio | rede-fio | 43 |
| rede computadores | rede-computadores | 41 |
| ad hoc | ad-hoc | 39 |

Tabela 19 – Exemplos de bigramas substituídos nos documentos do currículo *c*.

Fonte: Elaborado pelo autor.

Para o currículo *c*, a divisão dos 6 tópicos e as palavras com maior probabilidade de ocorrer em cada um são exibidas no mapa de calor da Figura 20. Percebe-se a correlação forte de todos os tópicos com a palavra *rede*, principalmente envolvendo *gerência*. As relações passam por *computação* em grade nos tópicos 1 e 3, ambos com viés de segurança; *gerência* de redes (tópicos 3, 5 e 6), sendo o tópico 5 com viés de sustentabilidade em nuvens-verdes; *nuvem* (tópicos 2 e 5); *gerência* com viés de monitoramento e qualidade de serviço ligadas a redes sem fio ad hoc (tópico 4). É visível que *segurança* (relacionada também com *deteção-intrusão*) permeia vários tópicos, muito provável pela questão da *gerência* estar fortemente presente (que indica, entre outras coisas, cuidado com a segurança). Na página do Currículo Lattes, existe o seguinte trecho informado pelo próprio autor: *Tem experiência na área de Ciência da Computação e Telecomunicações, com ênfase em Administração e Gerência de Redes e Serviços, atuando principalmente nos seguintes temas: segurança, computação autônoma, computação em nuvem e internet das coisas*. Alguns exemplos de documentos e seus respectivos anos podem ser conferidos na Tabela 20.

| Ano | Documento (parcial) |
|------|---|
| 2004 | Análise de Parâmetros de QoS e Esquemas de Filas em Ambiente Bluetooth |
| 2006 | Defending Grids Against Intrusions |
| 2011 | Uma abordagem para alocação e distribuição de máquinas virtuais em ambientes de computação em nuvem verde |
| 2012 | Aperfeiçoando a Gerência de Recursos para Nuvens Verdes |
| 2012 | Gerência, Segurança e Sustentabilidade para Computação em Nuvem |
| 2014 | Projeto GESSUCON ? GERência, Segurança e SUStentabilidade para COMpu-tação em Nuvem |
| 2018 | Projeto GAS CoNuInCo - Gerência Autônoma e Segurança para Computação em Nuvem e Internet das Coisas |
| 2018 | Gerência Autônoma para Computação em Nuvem |
| 2018 | Questões de Segurança e Privacidade dentro da Computação em Nuvem |

Tabela 20 – Exemplos de produções extraídas diretamente da página do currículo *c*

Fonte: Adaptado do Currículo Lattes.

Sobre a temporalidade, os tópicos 4 e 6 estão levemente deslocados no tempo. Avaliando a composição de palavras, um parece a evolução do outro, ou, minimamente, uma variação. Os tópicos mais proeminentes são o 2 e 5, que envolvem justamente o conceito de *cloud computing*, um com mais enfoque em segurança e outro em sustentabilidade. O gráfico de probabilidades acumuladas da Figura 22 mostra que os tópicos 1, 3, 4 e 6 coexistiram por algum tempo, mas os esforços nos últimos anos estão voltados para os tópicos 2 e 5.

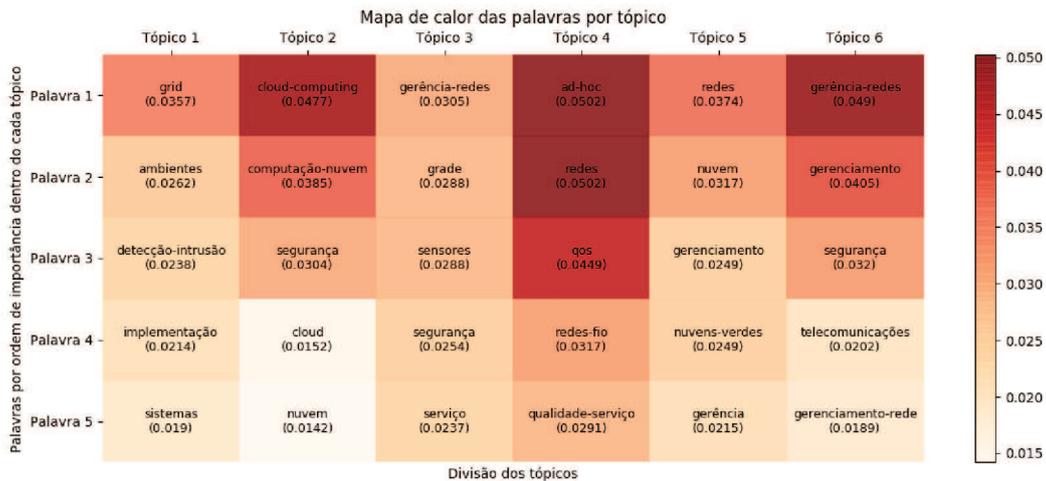


Figura 20 – Mapa de calor das palavras por tópico do currículo c
Fonte: Elaborado pelo autor.

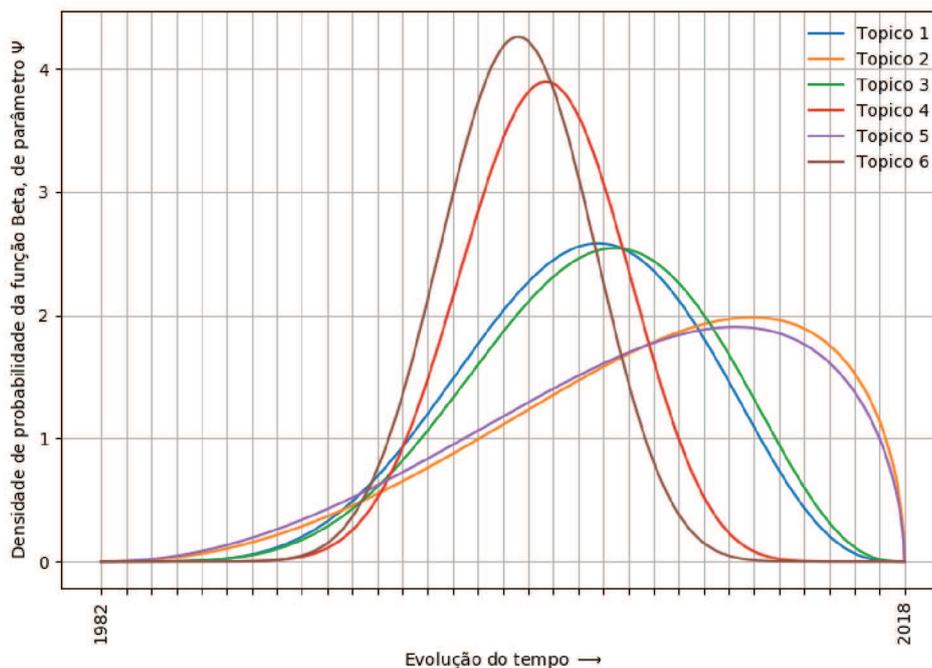


Figura 21 – Evolução dos tópicos ao longo do tempo do currículo c
Fonte: Elaborado pelo autor.

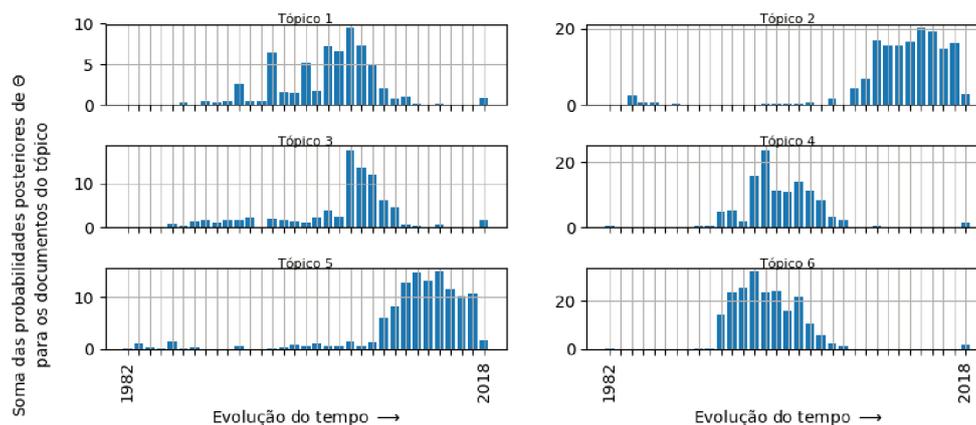


Figura 22 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo c

Fonte: Elaborado pelo autor.

5.4.2 Experimento para grupo de pesquisadores

Para a execução com todos os currículos do PPGCC, as principais palavras de cada tópico podem ser conferidas na Figura 23. Para fazer uma análise com esses tópicos, foram utilizadas as linhas de pesquisa do Programa e suas subdivisões, conforme a Tabela 21.

| Linhas de pesquisa | Subdivisão |
|---|--|
| 1. Computação Paralela e Distribuída | <ul style="list-style-type: none"> a) Computação em Nuvem b) Computação de Alto Desempenho c) Web Services d) Tolerância a Falhas e) Sistemas de Tempo Real f) Internet das Coisas g) Paralelização de Algoritmos |
| 2. Banco de Dados | <ul style="list-style-type: none"> a) Big Data b) Mineração de Dados c) Data Science d) Recuperação de Informação e) Gerenciamento de Dados Complexos f) Processamento de Linguagem Natural g) Web Semântica |
| 3. Engenharia de Software | <ul style="list-style-type: none"> a) Gerência de Projetos b) Computação na Escola c) Métodos Ágeis d) Modelagem de Dados e) Métodos Formais Aplicados f) Interação Humana-Computador |
| 4. Inteligência Computacional | <ul style="list-style-type: none"> a) Computação Gráfica b) Reconhecimento de Padrões c) Aprendizado de Máquina d) Sistemas Multiagentes e) Robótica f) Agentes Inteligentes g) Inteligência Artificial |
| 5. Redes de Computadores | <ul style="list-style-type: none"> a) Sistemas Distribuídos b) Protocolos de Comunicação c) Redes de Sensores d) Internet das Coisas e) Computação em Nuvem |
| 6. Segurança em Sistemas Computacionais | <ul style="list-style-type: none"> a) Segurança Cibernética b) Segurança em Sistemas Distribuídos c) Criptografia d) Certificação Digital e) Cripto-moedas f) Blockchain g) Privacidade |
| 7. Sistemas Embarcados | <ul style="list-style-type: none"> a) Redes de Sensores b) Internet das Coisas c) Microeletrônica d) Processamento de Sinais Digitais e) Automação de Projeto Eletrônico |

Tabela 21 – Linhas de pesquisa do PPGCC e suas subdivisões.

Fonte: Elaborado pelo autor.

Através da Tabela 21 acima, durante a análise cada um dos tópicos é válido destacar os seguintes pontos:

- Sobre o tópico 1, pode-se dizer que está relacionado exclusivamente com a linha 3, tratando de processos de software, que envolvem planejamento, melhoria, qualidade, sendo o CMMI um exemplo de processo, tendo a ISO-IEC definido vários outros processos;
- Sobre o tópico 2, há uma semelhança muito grande com os elementos da linha 2, principalmente com o 2.f - Processamento de Linguagem Natural e 2.g - Web semântica, passando também por outros elementos da mesma linha;
- No tópico 3, há uma aderência com a linha de pesquisa 4, passando pelo 4.b - Reconhecimento de Padrões (que envolve o 4.c - Aprendizado de Máquina), 4.d - Sistemas Multi Agentes e 4.f - Agentes Inteligentes, que são todos itens da 4.g - Inteligência Artificial;
- Com relação ao tópico 5, parece haver relação com a linha 3, com mais proximidade ao 3.b - Computação na escola e com 3.f - Interação Humana-Computador;
- No tópico 6, há uma forte relação com a linha de pesquisa 4, envolvendo os itens 4.1, 4.b, 4.c e 4.d (e inerentemente o 4.g). Os assuntos desse tópico (imagens e telemedicina) são abordados por pesquisadores do departamento de ensino onde a maioria dos professores do PPGCC estão lotados;
- O tópico 7 parece ter sido isolado para tratar do item 3.b. Aqui aparecem algumas palavras que sugerem o assunto do tópico 6. Isso provavelmente deve ter acontecido pela proximidade dos pesquisadores que tratam dos dois assuntos;
- O tópico 8 tem forte relação com a linha 6, mais especificamente com 6.a e 6.d, também com a linha 5, e uma mistura da linha 1, que podem ser considerados correlacionados;
- Os dois últimos tópicos são os que apresentam as misturas, tornando-os mais abrangentes e menos específicos. No caso do tópico 9, há presença das linhas de pesquisa 2, 3, 4 e 5 (misturando banco de dados, com redes, inteligência artificial, web e engenharia de software), e no tópico 10 a presença das linhas de pesquisa 3 (design, computação na escola, modelagem), 4 (sistemas multiagentes) e 7 (microeletrônica, timing, circuitos integrados, systemc).

Pela análise acima, é possível perceber que todas as linhas de pesquisa foram contempladas de alguma forma, com destaque para a área 3 - Engenharia de Software e 4 -

Inteligência Computacional, que permeiam vários tópicos. Vale lembrar que elas podem ter sido potencializadas pelas demais áreas, uma vez que são ortogonais. Inteligência Artificial, por exemplo, pode ser utilizada em redes de computadores para detecção de intrusão, em banco de dados para data mining, em sistemas embarcados para melhoria de projetos digitais, na nuvem para detecção de falhas, entre outros. Engenharia de Software pode estar presente na modelagem de dados, melhoria de processos, análise de requisitos, gestão, usabilidade, verificação, validação e testes de qualquer uma das outras linhas de pesquisa (inclusive de sistemas embarcados).

Analisando a temporalidade nos gráficos das Figuras 24 e 25, inicialmente se destacam justamente os dois tópicos que apresentam maior mistura de linhas de pesquisa (9 e 10). Eles aparecem isolados, aproximadamente na metade da vida do programa. A mistura deve ter acontecido por algum padrão de coocorrência das palavras nessa época, que não se repetiu nos anos seguintes. Os demais tópicos parecem seguir a mesma tendência ao longo do tempo, com produção relativamente constante entre eles. Um destaque é para o assunto do tópico 7, que parece ter surgido no início do 4^o quartil do espaço de tempo, tendo produções cada vez maiores, assim como uma diminuição gradativa da produção no tópico 6. Por outro lado, o tópico 3 (que parece tratar de reconhecimento de padrões, assim como o 6), está em alta nos últimos anos. Neste caso, pode ser adicionada a interpretação de que está ocorrendo uma mudança no foco desse tópico, onde os assuntos das publicações estão sendo alteradas gradativamente.

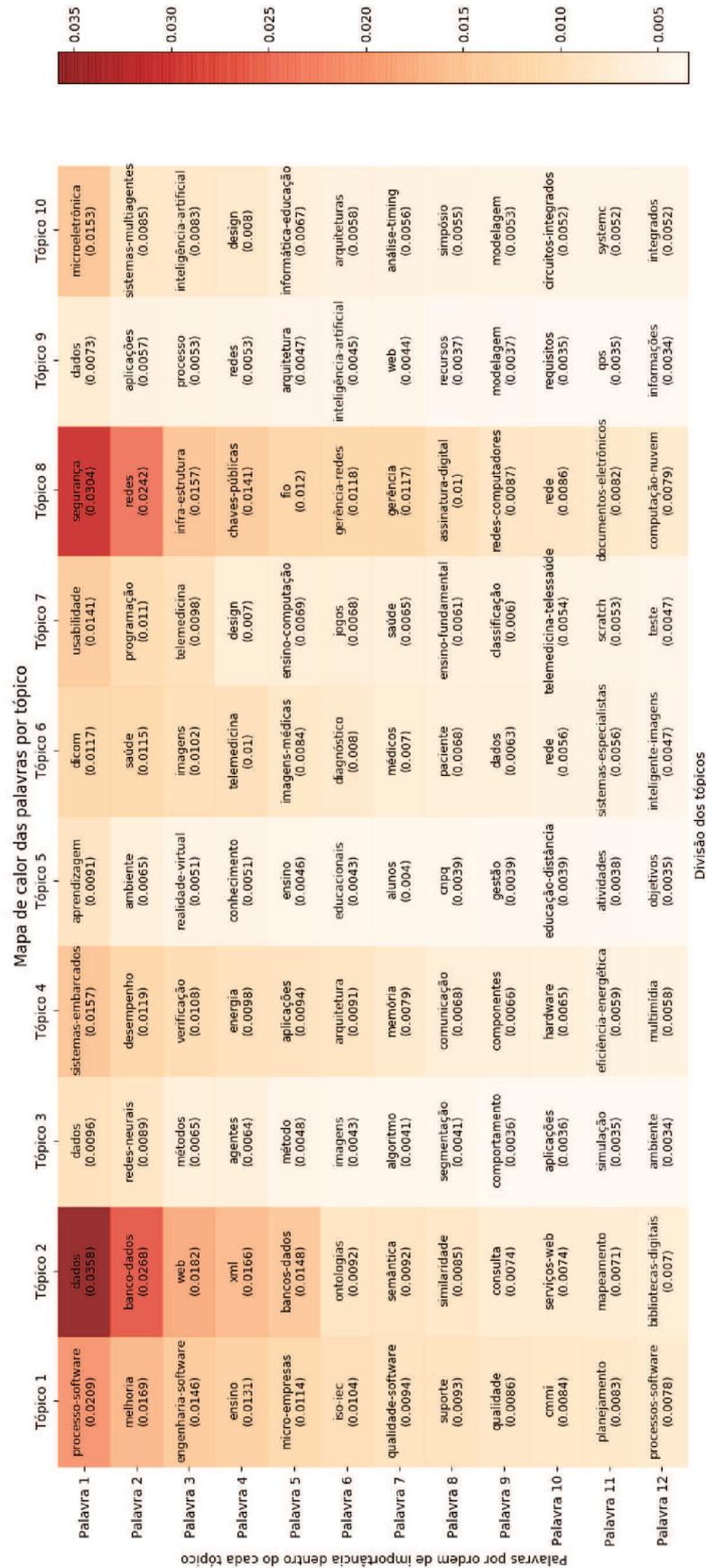


Figura 23 – Mapa de calor das palavras por tópico de todos os currículos do PPGCC
 Fonte: Elaborado pelo autor.

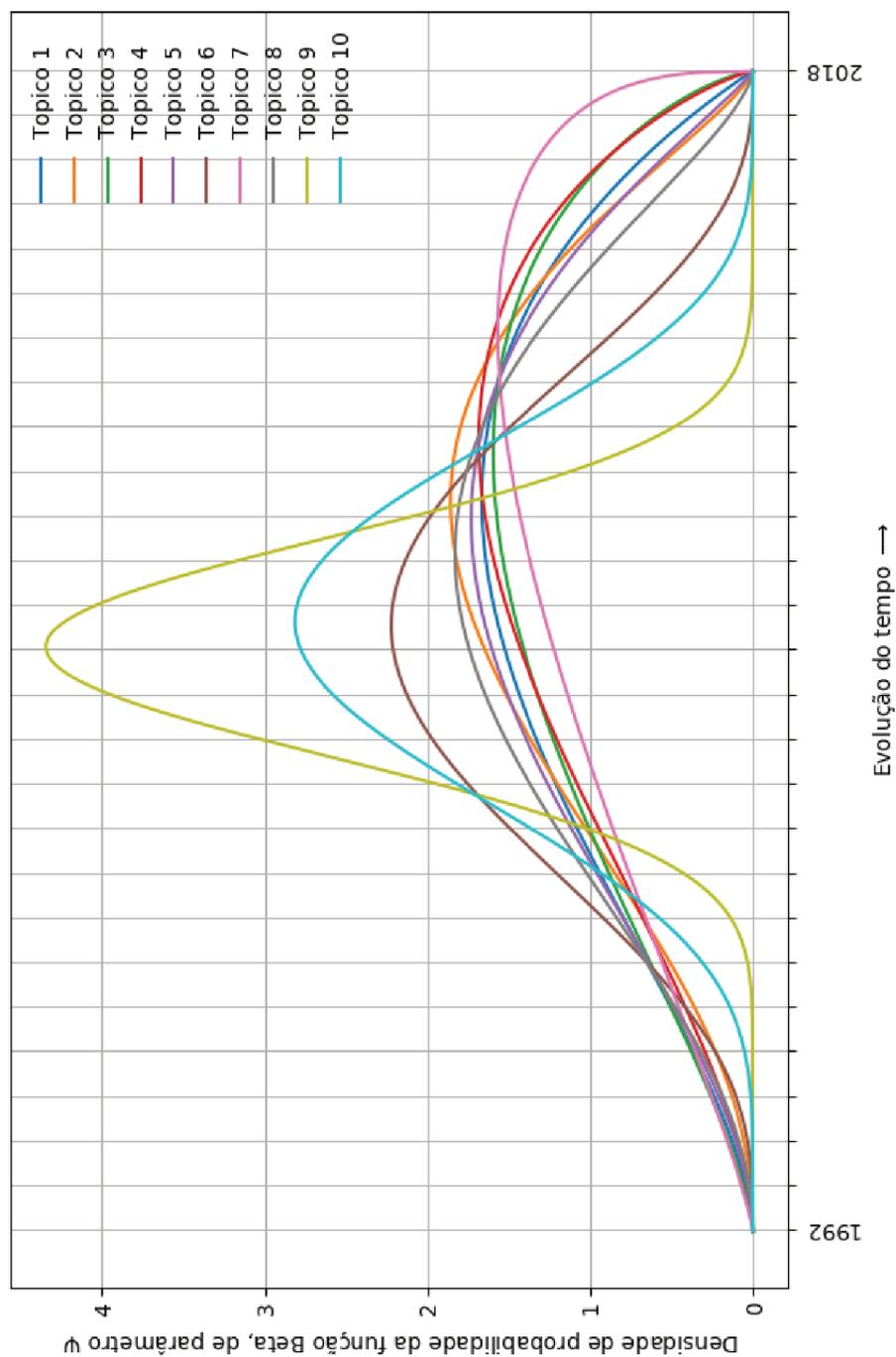


Figura 24 – Evolução dos tópicos ao longo do tempo de todos os currículos do *PPGCC*
Fonte: Elaborado pelo autor.

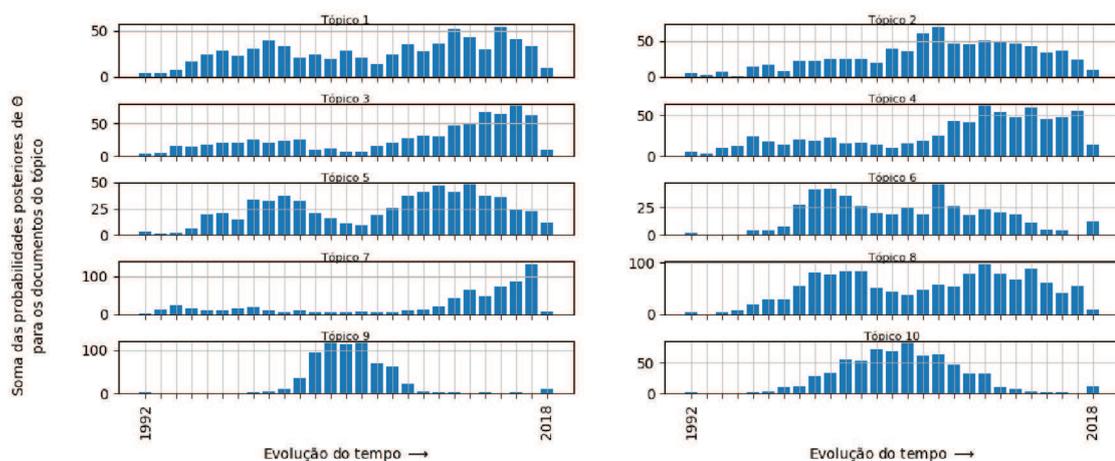


Figura 25 – Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, de todos os currículos do *PPGCC*

Fonte: Elaborado pelo autor.

6 Considerações Finais

6.1 Limitações

A primeira limitação conhecida é relacionada à fonte de dados, ou seja, os currículos da Plataforma Lattes. Mesmo que exista uma cobrança, tanto dos órgãos avaliadores do Ensino Superior, como da própria legislação, muitos pesquisadores não mantêm as informações atualizadas, principalmente das últimas publicações. Isso gera um vazio de registros, justamente no período que se deseja analisar (geralmente as últimas publicações são as que indicam em que área o pesquisador está atuando no momento). Além disso, as informações disponibilizadas são resumidas. Basicamente o que está disponível é o título dos artigos e demais trabalhos, se resumindo a uma ou duas frases, com exceção de alguns projetos, que apresentam uma descrição maior. Vale ressaltar, ainda, que foi utilizada apenas uma fonte de dados e, apesar da sua extensão, seria interessante integrar com outras fontes, mas foge do escopo deste trabalho.

Ainda sobre a fonte de dados, todos os registros da produção do pesquisador são considerados com o mesmo grau de importância. Aqui devem ser feitas algumas discussões sobre a possível distorção dessa abordagem. Primeiro, o tipo de trabalho poderia ser considerado como influenciador no peso. Uma orientação de doutorado poderia ter um peso maior do que uma orientação de TCC, da mesma forma que um capítulo de livro sobre uma publicação em anais de eventos científicos. A classificação Qualis, apesar de todas as suas controvérsias, também poderia ser utilizada para aumentar o peso das publicações consideradas mais *importantes*, nos meios A1 e A2, diminuindo a importância gradativamente para as classificações B1, B2, B2, B4, B5 e C.

A influência sobre o posicionamento da autoria dos pesquisadores nas produções bibliográficas também não é considerada. Caso o pesquisador tenha um trabalho como 1º autor e outro onde aparece como 5º, ambos serão considerados como iguais pelo modelo (ou mesmo a influência de quando é orientador de algum trabalho acadêmico ou quando é o próprio autor). Um mecanismo para atenuar a posição poderia ser utilizado para equalizar os pesos, assim como na classificação Qualis.

Outro ponto que vale ser ressaltado é que os documentos utilizados no modelo possuem apenas o ano para indicar a marcação do tempo (e apenas um). Algumas produções demoram vários anos, como é o caso de projetos ou de orientações de doutorado (4 anos, em média). Neste trabalho, optou-se por utilizar o ano inicial do documento. Por exemplo: uma orientação de 2012 à 2015 é inserida no modelo com o ano de 2012. Nesta situação poderiam ser adotadas abordagens de replicação (replicar o mesmo conteúdo para os

anos de vigência do documento) ou ainda uma função que descreva sua importância em diferentes anos. Como o impacto da adoção dessas abordagens não foi estudado, nenhuma foi aplicada no projeto em questão.

Sobre o modelo usado no núcleo do framework, já foi discutido que o TOT trata os tópicos de maneira imutável. A imutabilidade dos tópicos pode ser interessante em alguns casos, mas em outros pode ser um fator limitante. Ter a evolução de um determinado tópico pode trazer suas diferentes composições ao longo do tempo (como ocorre no TET). No caso em que é tratado como imutável, uma pequena alteração na composição se desdobra no surgimento de um novo tópico (como ocorre no TOT). A função Beta utilizada dentro do TOT para representação do tempo é monomodal, portanto não é capaz de detectar algumas variações nos dados, que só são visíveis no gráfico de probabilidades acumuladas.

6.2 Trabalhos Futuros

De um modo geral, a melhoria de todas as limitações citadas acima podem ser consideradas para trabalhos futuros, como utilização de outras fontes de dados, funções de decaimento para diferentes tipos de documento, utilização de uma função bimodal para representação do tempo, melhoria na representação do tempo por documentos, entre outros.

O processamento do texto de entrada é peça fundamental para geração de bons resultados. Seria interessante realizar um estudo mais aprofundado de técnicas de limpeza de dados, assim os dados teriam mais relevância. Um passo de processamento de texto que não foi incorporado a este trabalho é a normalização morfológica: na Língua Portuguesa, as palavras podem variar de acordo com gênero (masculino ou feminino), número (singular ou plural), grau (aumentativo ou diminutivo), pessoa (primeira, segunda ou terceira pessoa do singular ou plural), modo (indicativo, subjuntivo e imperativo), tempo (presente, pretérito e futuro), voz (ativa, passiva e reflexiva), conforme a classificação gramatical, e ainda possuir prefixos ou sufixos. Essa característica faz com que palavras como *rede* e *redes* sejam diferentes, tendo menos importância individualmente. Caso elas fossem transformadas para sua raiz antes da análise, ambas poderiam ser tratadas apenas como *rede*.

A busca por especialistas que possuam similaridade entre si ou com um grupo também pode ser implementada utilizando a função de divergência de Kullback–Leibler (ou divergência KL). Por exemplo, ao aplicar o TOT para um pesquisador qualquer, sua similaridade pode ser comparada ao PPGCC através da divergência KL, para saber a proximidade deste pesquisador com as linhas do programa. A métrica KL pode ser aplicada entre resultados da execução do TOT.

Outro ponto que pode ser evoluído é o modelo utilizado. Como discutido nas seções iniciais, o TOT possui uma abordagem de tópicos imutáveis. Isso confere uma certa rigidez

e sensibilidade aos tópicos gerados (alterações pequenas geram novos tópicos). Existem modelos mais robustos, como o TAT, que são capazes de capturar também a mutação dos tópicos, quer permite que algumas palavras ganhem força em detrimento de outras ao longo do tempo, dentro do mesmo tópico.

Caso um modelo baseado em LDA continue sendo utilizado, pode ser realizado um estudo aprofundado nas técnicas para determinação do número ideal de tópicos, bem como parametrização adequada de α e β .

6.3 Conclusão

No início deste trabalho, foi feita uma motivação pertinente para determinação da especialidade de uma determinada pessoa levando em consideração o fator tempo. Com essa problemática, foi proposto o desenvolvimento de um framework capaz de determinar a especialidade ao longo do tempo de maneira não-supervisionada, usando como fonte de dados os currículos cadastrados na Plataforma Lattes, com posterior aplicação com os membros de Programa de Pós-Graduação da UFSC, sendo escolhido o PPGCC. Ao longo da fundamentação, foram discutidos vários conceitos importantes que formaram a base para o desenvolvimento da ferramenta proposta. Em resumo, o framework agrupa as palavras dos documentos analisados em tópicos, que por sua vez são relacionados com a especialidade da pessoa.

Para verificação do framework, foram feitos testes com 3 pesquisadores diferentes, e posterior aplicação no simultaneamente em 27 pesquisadores do PPGCC. Neste ponto fica evidente a influência do tempo, onde alguns tópicos emergem em detrimento de outros. Mais especificamente no caso do Programa de Pós Graduação, os tópicos de maior relação com as linhas de pesquisa permanecem mais constantes no tempo, como era esperado, pois precisam se manter (mesmo que um de seus integrantes mude o foco da pesquisa, a linha permanece com demais membros pesquisando na mesma área).

Duas das dificuldades foram (i) encontrar o valor ideal de tópicos e dos parâmetros α e β e (ii) efetuar um filtro correto nas palavras. Com relação ao filtro, a ação está condicionada na granularidade desejada por quem está analisando os dados: quanto mais específica a necessidade dos dados, mais as palavras genéricas precisam ser removidas dos documentos. No caso dos parâmetros, foram encontrados baseados em testes consecutivos, pela familiaridade com os dados analisados, ajustando-os pouco a pouco. Nada impede que os valores genéricos propostos na literatura sejam usados, mas um ajuste fino pode trazer uma grande melhora.

Por fim, a ferramenta apresentada mostra-se como uma opção para análise da temporalidade em conjuntos de dados, com uma possível inclusão nos processos de gestão de pesquisadores. Para que o modelo seja mais robusto e possa atender de forma efetiva e

real, os itens propostos nos trabalhos futuros devem ser considerados e tratados.

Referências

- ALVES, R. C. V. et al. Ciência da informação, ciência da computação e recuperação da informação: algumas considerações sobre os métodos e tecnologias da informação utilizados ao longo do tempo. *Revista Eletrônica Informação e Cognição*, v. 6, n. 1, p. 28–40, 2007. ISSN 1807-8281. Citado na página 13.
- BALOG, K.; FANG, Y.; RIJKE, M. de. *Expertise Retrieval*. [S.l.]: Now Publishers, 2012. (Foundations and Trends in Information Retrieval). ISBN 9781601985668. Citado 2 vezes nas páginas 13 e 14.
- BLEI, D. M. Probabilistic topic models. *Commun. ACM*, ACM, New York, NY, USA, v. 55, n. 4, p. 77–84, abr. 2012. ISSN 0001-0782. Disponível em: <<http://doi.acm.org/10.1145/2133806.2133826>>. Citado na página 24.
- CAMPBELL, C. S. et al. Expertise identification using email communications. In: *Proceedings of the Twelfth International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2003. (CIKM '03), p. 528–531. ISBN 1-58113-723-0. Disponível em: <<http://doi.acm.org/10.1145/956863.956965>>. Citado na página 25.
- CNPQ. *Lattes - Extração de Dados*. 2018. Disponível em: <<http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>>. Citado na página 17.
- CNPQ. *Lattes Histórico - História do Surgimento da Plataforma Lattes*. 2018. Disponível em: <<http://lattes.cnpq.br/web/plataforma-lattes/historico>>. Citado na página 17.
- CNPQ. *A Plataforma Lattes*. 2018. Disponível em: <<http://lattes.cnpq.br/web/plataforma-lattes/o-que-e>>. Citado na página 17.
- DAUD, A. Using time topic modeling for semantics-based dynamic research interest finding. *Knowledge-Based Systems*, v. 26, p. 154 – 163, 2012. ISSN 0950-7051. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705111001638>>. Citado 2 vezes nas páginas 33 e 34.
- DAUD, A. et al. Exploiting temporal authors interests via temporal-author-topic modeling. In: HUANG, R. et al. (Ed.). *Advanced Data Mining and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 435–443. ISBN 978-3-642-03348-3. Citado 2 vezes nas páginas 14 e 33.
- DAUD, A. et al. Temporal expert finding through generalized time topic modeling. *Know.-Based Syst.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 23, n. 6, p. 615–625, ago. 2010. ISSN 0950-7051. Citado 5 vezes nas páginas 14, 15, 23, 34 e 36.
- DOM, B. et al. Graph-based ranking algorithms for e-mail expertise analysis. In: *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. New York, NY, USA: ACM, 2003. (DMKD '03), p. 42–48. Disponível em: <<http://doi.acm.org/10.1145/882082.882093>>. Citado na página 24.

FALEIROS, T. D. P.; LOPES, A. D. A. *Modelos Probabilísticos de Tópicos: desvendando o Latent Dirichlet Allocation*. [S.l.], 2016. Citado 4 vezes nas páginas 24, 28, 29 e 30.

FANG, Y.; GODAVARTHY, A. Modeling the dynamics of personal expertise. In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. New York, NY, USA: ACM, 2014. (SIGIR '14), p. 1107–1110. ISBN 978-1-4503-2257-7. Disponível em: <<http://doi.acm.org/10.1145/2600428.2609521>>. Citado na página 30.

GIORDANO, R.; BIOLCHINI, J. Busca e recuperação da informação científica na web: comportamento informacional de profissionais da informação. *InCID: Revista de Ciência da Informação e Documentação*, v. 3, n. 1, p. 125–145, 2012. ISSN 2178-2075. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42374>>. Citado na página 13.

LI, Y.; TANG, J. Expertise search in a time-varying social network. In: *2008 The Ninth International Conference on Web-Age Information Management*. [S.l.: s.n.], 2008. p. 293–300. Citado 3 vezes nas páginas 14, 36 e 38.

LIN, S. et al. A survey on expert finding techniques. *Journal of Intelligent Information Systems*, *Journal of Intelligent Information Systems*, v. 49, n. 2, p. 255–279, 2017. ISSN 15737675. Citado 3 vezes nas páginas 22, 24 e 27.

MACDONALD, C. The voting model for people search. *ACM SIGIR Forum*, v. 43, 06 2009. Citado 2 vezes nas páginas 25 e 26.

MATA-RIVERA, F. et al. A collaborative learning approach for geographic information retrieval based on social networks. *Computers in Human Behavior*, v. 51, p. 829 – 842, 2015. ISSN 0747-5632. Citado na página 13.

MAURYA, A. Topics over time - an open-source implementation. 2015. Citado na página 53.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, v. 2, n. 1, p. 20–32, 1951. Citado na página 13.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. *Mineração de Textos*. [S.l.], 2007. 29 p. Citado na página 51.

NGUYEN, D. Q. jLDADMM: A Java package for the LDA and DMM topic models. *arXiv preprint arXiv:1808.03835*, 2018. Citado na página 53.

ROSEN-ZVI, M. et al. Learning author-topic models from text corpora. *ACM Transactions on Information Systems*, v. 28, n. 1, p. 1–38, 2010. ISSN 10468188. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1658377.1658381>>. Citado na página 24.

RUSSELL-ROSE, T.; CHAMBERLAIN, J. Searching for talent: The information retrieval challenges of recruitment professionals. *Business Information Review*, v. 33, n. 1, p. 40–48, 2016. Citado na página 14.

RYBAK, J.; BALOG, K.; NØRVÅG, K. Temporal expertise profiling. In: *Proceedings of the 36th European Conference on IR Research on Advances in Information Retrieval - Volume 8416*. New York, NY, USA: Springer-Verlag New York, Inc., 2014. (ECIR 2014), p. 540–546. ISBN 978-3-319-06027-9. Citado 4 vezes nas páginas 14, 31, 32 e 33.

- SORNIL, O. *A Graph-Based Information Retrieval Model*. 2010. Citado na página 24.
- STEYVERS, M.; GRIFFITHS, T. Latent semantic analysis: A road to meaning. In: _____. [S.l.]: Laurence Erlbaum, 2007. cap. Probabilistic topic models. Citado na página 28.
- TALABIS, M. R. M. et al. Chapter 1 - analytics defined. In: TALABIS, M. R. M. et al. (Ed.). *Information Security Analytics*. Boston: Syngress, 2015. p. 1 – 12. ISBN 978-0-12-800207-0. Citado na página 27.
- WANG, X.; MCCALLUM, A. Topics over time: A non-markov continuous-time model of topical trends. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2006. (KDD '06), p. 424–433. ISBN 1-59593-339-5. Disponível em: <<http://doi.acm.org/10.1145/1150402.1150450>>. Citado 3 vezes nas páginas 38, 40 e 55.

Apêndices

APÊNDICE A – Estrutura do Currículo Lattes

A partir da análise da estrutura do Currículo Lattes na versão XML, a lista abaixo foi considerada adequada para extração de documentos relevantes na determinação da expertise de cada pesquisador. A estrutura foi convertida em formato JSON, para ser importada pelo framework.

```

1  [
2  {
3    "descricao" : "Formacao academica - Graduacoes",
4    "tag" : ".//FORMACAO-ACADEMICA-TITULACAO/GRADUACAO",
5    "inicio" : ["./@ANO-DE-INICIO"],
6    "fim" : ["./@ANO-DE-CONCLUSAO"],
7    "atributos" : ["./@TITULO-DO-TRABALHO-DE-CONCLUSAO-DE-CURSO"]
8  },
9  {
10   "descricao" : "Formacao academica - Mestrado",
11   "tag" : ".//FORMACAO-ACADEMICA-TITULACAO/MESTRADO",
12   "inicio" : ["./@ANO-DE-INICIO"],
13   "fim" : ["./@ANO-DE-OBTENCAO-DO-TITULO", "./@ANO-DE-CONCLUSAO"],
14   "atributos" : [
15     "./@TITULO-DA-DISSERTACAO-TESE",
16     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
17     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
18     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
19     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
20     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
21     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"
22   ]
23 },
24 {
25   "descricao" : "Formacao academica - Doutorado",
26   "tag" : ".//FORMACAO-ACADEMICA-TITULACAO/DOCTORADO",
27   "inicio" : ["./@ANO-DE-INICIO"],
28   "fim" : ["./@ANO-DE-OBTENCAO-DO-TITULO", "./@ANO-DE-CONCLUSAO"],
29   "atributos" : [
30     "./@TITULO-DA-DISSERTACAO-TESE",
31     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
32     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
33     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
34     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
35     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
36     ".//PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"
37   ]
38 },
39 {
40   "descricao" : "Atuacoes profissionais (apenas vinculos + outras
41     informacoes)",
42   "tag" : ".//ATUACOES-PROFISSIONAIS/ATUACAO-PROFISSIONAL/VINCULOS",
43   "inicio" : ["./@ANO-INICIO"],
44   "fim" : ["./@ANO-FIM"],
45   "atributos" : ["./@OUTRAS-INFORMACOES"]
46 },
47 {
48   "descricao" : "Atuacoes profissionais (Atividades de administracao) "
49   "tag" : ".//ATUACOES-PROFISSIONAIS/ATUACAO-PROFISSIONAL/ATIVIDADES-
50     DE-DIRECAO-E-ADMINISTRACAO/DIRECAO-E-ADMINISTRACAO",
51   "inicio" : ["./@ANO-INICIO"],
52   "fim" : ["./@ANO-FIM"],

```

```
51     "atributos" : [ "@CARGO-OU-FUNCAO" ]
52   },
53   {
54     "descricao" : "Pesquisa e desenvolvimento (linhas de pesquisa)",
55     "tag" : "://ATUACOES-PROFISSIONAIS/ATUACAO-PROFISSIONAL/ATIVIDADES-
        DE-PESQUISA-E-DESENVOLVIMENTO/PESQUISA-E-DESENVOLVIMENTO/LINHA-DE-
        PESQUISA",
56     "inicio" : [ "@ANO-INICIO" ],
57     "fim" : [ "@ANO-FIM" ],
58     "atributos" : [
59       "@TITULO-DA-LINHA-DE-PESQUISA",
60       "@OBJETIVOS-LINHA-DE-PESQUISA",
61       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
62       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
63       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
64       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
65       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
66       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"
67     ]
68   },
69   {
70     "descricao" : "Atividades de extensao universitaria",
71     "tag" : "://ATUACOES-PROFISSIONAIS/ATUACAO-PROFISSIONAL/ATIVIDADES-
        DE-EXTENSAO-UNIVERSITARIA/EXTENSAO-UNIVERSITARIA",
72     "inicio" : [ "@ANO-INICIO" ],
73     "fim" : [ "@ANO-FIM" ],
74     "atributos" : [
75       "@ATIVIDADE-DE-EXTENSAO-REALIZADA"
76     ]
77   },
78   {
79     "descricao" : "Participacoes em projetos de DESENVOLVIMENTO, EXTENSAO
        , PESQUISA, OUTRA",
80     "tag" : "://ATUACOES-PROFISSIONAIS/ATUACAO-PROFISSIONAL/ATIVIDADES-
        DE-PARTICIPACAO-EM-PROJETO/PARTICIPACAO-EM-PROJETO/PROJETO-DE-
        PESQUISA",
81     "inicio" : [ "@ANO-INICIO" ],
82     "fim" : [ "@ANO-FIM" ],
83     "atributos" : [
84       "@NOME-DO-PROJETO",
85       "@DESCRICAO-DO-PROJETO"
86     ]
87   },
88   {
89     "descricao" : "Trabalhos em eventos",
90     "tag" : "://PRODUCAO-BIBLIOGRAFICA/TRABALHOS-EM-EVENTOS/TRABALHO-EM-
        EVENTOS",
91     "inicio" : [ "DADOS-BASICOS-DO-TRABALHO/@ANO-DO-TRABALHO" ],
92     "fim" : [ "DADOS-BASICOS-DO-TRABALHO/@ANO-DO-TRABALHO" ],
93     "atributos" : [
94       "DADOS-BASICOS-DO-TRABALHO/@TITULO-DO-TRABALHO",
95       "DADOS-BASICOS-DO-TRABALHO/@TITULO-DOS-ANAIS-OU-PROCEEDINGS",
96       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
97       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
98       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
99       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
100      "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
101      "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
102      "INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
103     ]
104   },
105   {
106     "descricao" : "Artigos publicados",
107     "tag" : "://PRODUCAO-BIBLIOGRAFICA/ARTIGOS-PUBLICADOS/ARTIGO-
        PUBLICADO",
108     "inicio" : [ "DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO" ],
109     "fim" : [ "DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO" ],
110     "atributos" : [
111       "DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO",
112       "DETALHAMENTO-DO-ARTIGO/@TITULO-DO-PERIODICO-OU-REVISTA",
113       "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
114     ]
115   }
116 }
```

```

115     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
116     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
117     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
118     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
119     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
120     "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
121 ]
122 },
123 {
124     "descricao" : "Livros e capitulos (publicados ou organizados)",
125     "tag" : "./PRODUCAO-BIBLIOGRAFICA/LIVROS-E-CAPITULOS/LIVROS-
126     PUBLICADOS-OU-ORGANIZADOS/LIVRO-PUBLICADO-OU-ORGANIZADO",
127     "inicio" : ["/DADOS-BASICOS-DO-LIVRO/@ANO"],
128     "fim" : ["/DADOS-BASICOS-DO-LIVRO/@ANO"],
129     "atributos" : [
130         "/DADOS-BASICOS-DO-LIVRO/@TITULO-DO-LIVRO",
131         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
132         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
133         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
134         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
135         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
136         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
137         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
138     ]
139 },
140 {
141     "descricao" : "Capitulos de livros publicados",
142     "tag" : "./PRODUCAO-BIBLIOGRAFICA/LIVROS-E-CAPITULOS/CAPITULOS-DE-
143     LIVROS-PUBLICADOS/CAPITULO-DE-LIVRO-PUBLICADO",
144     "inicio" : ["/DADOS-BASICOS-DO-CAPITULO/@ANO"],
145     "fim" : ["/DADOS-BASICOS-DO-CAPITULO/@ANO"],
146     "atributos" : [
147         "/DADOS-BASICOS-DO-CAPITULO/@TITULO-DO-CAPITULO-DO-LIVRO",
148         "/DETALHAMENTO-DO-CAPITULO/TITULO-DO-LIVRO",
149         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
150         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
151         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
152         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
153         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
154         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
155         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
156     ]
157 },
158 {
159     "descricao" : "Texto em jornais ou revistas",
160     "tag" : "./PRODUCAO-BIBLIOGRAFICA/TEXTOS-EM-JORNAIS-OU-REVISTAS/
161     TEXTO-EM-JORNAL-OU-REVISTA",
162     "inicio" : ["/DADOS-BASICOS-DO-TEXTO/@ANO-DO-TEXTO"],
163     "fim" : ["/DADOS-BASICOS-DO-TEXTO/@ANO-DO-TEXTO"],
164     "atributos" : [
165         "/DADOS-BASICOS-DO-TEXTO/TITULO-DO-TEXTO",
166         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
167         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
168         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
169         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
170         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
171         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
172         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
173     ]
174 },
175 {
176     "descricao" : "Demais tipos de producao bibliografica",
177     "tag" : "./PRODUCAO-BIBLIOGRAFICA/DEMAIS-TIPOS-DE-PRODUCAO-
178     BIBLIOGRAFICA/OUTRA-PRODUCAO-BIBLIOGRAFICA",
179     "inicio" : ["/DADOS-BASICOS-DE-OUTRA-PRODUCAO/@ANO"],
180     "fim" : ["/DADOS-BASICOS-DE-OUTRA-PRODUCAO/@ANO"],
181     "atributos" : [
182         "/DADOS-BASICOS-DE-OUTRA-PRODUCAO/@TITULO",
183         "/DADOS-BASICOS-DE-OUTRA-PRODUCAO/@NATUREZA",
184         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
185         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",

```

```
182     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
183     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
184     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
185     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
186     "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
187 ]
188 },
189 {
190     "descricao" : "Artigos aceitos para publicacao",
191     "tag" : "./PRODUCAO-BIBLIOGRAFICA/ARTIGOS-ACEITOS-PARA-PUBLICACAO/
192     ARTIGO-ACEITO-PARA-PUBLICACAO",
193     "inicio" : [ "./DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO" ],
194     "fim" : [ "./DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO" ],
195     "atributos" : [
196         "./DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO",
197         "./DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-PERIODICO-OU-REVISTA",
198         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
199         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
200         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
201         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
202         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
203         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
204         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
205     ]
206 },
207 {
208     "descricao" : "Patentes",
209     "tag" : "./PRODUCAO-TECNICA/PATENTE",
210     "inicio" : [ "./DADOS-BASICOS-DA-PATENTE/@ANO-DESENVOLVIMENTO" ],
211     "fim" : [ "./DADOS-BASICOS-DA-PATENTE/@ANO-DESENVOLVIMENTO" ],
212     "atributos" : [
213         "./DADOS-BASICOS-DA-PATENTE/@TITULO",
214         "./DETALHAMENTO-DA-PATENTE/@FINALIDADE",
215         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
216         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
217         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
218         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
219         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
220         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
221         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
222     ]
223 },
224 {
225     "descricao" : "Produtos tecnologicos",
226     "tag" : "./PRODUCAO-TECNICA/PRODUTO-TECNOLOGICO",
227     "inicio" : [ "./DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO/@ANO" ],
228     "fim" : [ "./DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO/@ANO" ],
229     "atributos" : [
230         "./DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO/@TITULO-DO-PRODUTO",
231         "./DETALHAMENTO-DO-PRODUTO-TECNOLOGICO/@FINALIDADE",
232         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
233         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
234         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
235         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
236         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
237         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
238         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
239     ]
240 },
241 {
242     "descricao" : "Processos ou tecnicas",
243     "tag" : "./PRODUCAO-TECNICA/PROCESSOS-OU-TECNICAS",
244     "inicio" : [ "./DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS/@ANO" ],
245     "fim" : [ "./DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS/@ANO" ],
246     "atributos" : [
247         "./DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS/@TITULO-DO-PROCESSO",
248         "./DETALHAMENTO-DO-PROCESSOS-OU-TECNICAS/@FINALIDADE",
249         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
250         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
251         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
252         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
```

```

252     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
253     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
254     "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
255 ]
256 },
257
258 {
259     "descricao" : "Trabalhos tecnicos",
260     "tag" : "./PRODUCAO-TECNICA/TRABALHO-TECNICO",
261     "inicio" : ["/DADOS-BASICOS-DO-TRABALHO-TECNICO/@ANO"],
262     "fim" : ["/DADOS-BASICOS-DO-TRABALHO-TECNICO/@ANO"],
263     "atributos" : [
264         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
265         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
266         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
267         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
268         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
269         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
270         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
271     ]
272 },
273
274 {
275     "descricao" : "Apresentacao de trabalhos",
276     "tag" : "./PRODUCAO-TECNICA/DEMAIS-TIPOS-DE-PRODUCAO-TECNICA/
277     APRESENTACAO-DE-TRABALHO",
278     "inicio" : ["/DADOS-BASICOS-DA-APRESENTACAO-DE-TRABALHO/@ANO"],
279     "fim" : ["/DADOS-BASICOS-DA-APRESENTACAO-DE-TRABALHO/@ANO"],
280     "atributos" : [
281         "./DADOS-BASICOS-DA-APRESENTACAO-DE-TRABALHO/@TITULO",
282         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
283         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
284         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
285         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
286         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
287         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
288         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
289     ]
290 },
291 {
292     "descricao" : "Desenvolvimento de material didatico ou instrucional",
293     "tag" : "./PRODUCAO-TECNICA/DEMAIS-TIPOS-DE-PRODUCAO-TECNICA/
294     DESENVOLVIMENTO-DE-MATERIAL-DIDATICO-OU-INSTRUCIONAL",
295     "inicio" : ["/DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL/
296     @ANO"],
297     "fim" : ["/DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL/@ANO"],
298     "atributos" : [
299         "./DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL/@NATUREZA",
300         "./DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL/@TITULO",
301         "./DETALHAMENTO-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL/@FINALIDADE"
302     ],
303     "atributos" : [
304         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
305         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
306         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
307         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
308         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
309         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
310         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
311     ]
312 },
313 {
314     "descricao" : "Relatorio de pesquisa",
315     "tag" : "./PRODUCAO-TECNICA/DEMAIS-TIPOS-DE-PRODUCAO-TECNICA/
316     RELATORIO-DE-PESQUISA",
317     "inicio" : ["/DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA/@ANO"],
318     "fim" : ["/DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA/@ANO"],
319     "atributos" : [
320         "./DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA/@TITULO",
321         "./DETALHAMENTO-DO-RELATORIO-DE-PESQUISA/@NOME-DO-PROJETO",
322         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
323         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",

```

```
318     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
319     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
320     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
321     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
322     "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
323 ]
324 },
325 {
326     "descricao" : "Outra producao tecnica",
327     "tag" : "./PRODUCAO-TECNICA/DEMAIS-TIPOS-DE-PRODUCAO-TECNICA/OUTRA-
328     PRODUCAO-TECNICA",
329     "inicio" : ["/DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA/@ANO"],
330     "fim" : ["/DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA/@ANO"],
331     "atributos" : [
332         "./DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA/@NATUREZA",
333         "./DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA/@TITULO",
334         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
335         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
336         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
337         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
338         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
339         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
340         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
341     ]
342 },
343 {
344     "descricao" : "Orientacoes concluidas - Mestrado",
345     "tag" : "./OUTRA-PRODUCAO/ORIENTACOES-CONCLUIDAS/ORIENTACOES-
346     CONCLUIDAS-PARA-MESTRADO",
347     "inicio" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO/
348     @ANO"],
349     "fim" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO/
350     @ANO"],
351     "atributos" : [
352         "./DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO/@TITULO"
353         ,
354         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
355         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
356         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
357         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
358         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
359         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
360         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
361     ]
362 },
363 {
364     "descricao" : "Orientacoes concluidas - Doutorado",
365     "tag" : "./OUTRA-PRODUCAO/ORIENTACOES-CONCLUIDAS/ORIENTACOES-
366     CONCLUIDAS-PARA-DOCTORADO",
367     "inicio" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO
368     /@ANO"],
369     "fim" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO/
370     @ANO"],
371     "atributos" : [
372         "./DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO/@TITULO"
373         ,
374         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
375         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
376         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
377         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
378         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
379         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
380         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
381     ]
382 },
383 {
384     "descricao" : "Orientacoes concluidas - Pos-Doutorado",
385     "tag" : "./OUTRA-PRODUCAO/ORIENTACOES-CONCLUIDAS/ORIENTACOES-
386     CONCLUIDAS-PARA-POS-DOCTORADO",
387     "inicio" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-
388     DOCTORADO/@ANO"],
```

```
379 "fim" : ["/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-
      DOUTORADO/@ANO"],
380 "atributos" : [
381   "/DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-DOUTORADO/
      @TITULO",
382   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
383   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
384   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
385   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
386   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
387   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
388   "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
389 ]
390 },
391 {
392 "descricao" : "Orientacoes concluidas - Outras",
393 "tag" : "/OUTRA-PRODUCAO/ORIENTACOES-CONCLUIDAS/OUTRAS-ORIENTACOES-
      CONCLUIDAS",
394 "inicio" : ["/DADOS-BASICOS-DE-OUTRAS-ORIENTACOES-CONCLUIDAS/@ANO"],
395 "fim" : ["/DADOS-BASICOS-DE-OUTRAS-ORIENTACOES-CONCLUIDAS/@ANO"],
396 "atributos" : [
397   "/DADOS-BASICOS-DE-OUTRAS-ORIENTACOES-CONCLUIDAS/@TITULO",
398   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
399   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
400   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
401   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
402   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
403   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
404   "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
405 ]
406 },
407 {
408 "descricao" : "Participacao em congressos",
409 "tag" : "/DADOS-COMPLEMENTARES/PARTICIPACAO-EM-EVENTOS-CONGRESSOS/
      PARTICIPACAO-EM-CONGRESSO",
410 "inicio" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-CONGRESSO/@ANO"],
411 "fim" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-CONGRESSO/@ANO"],
412 "atributos" : [
413   "/DADOS-BASICOS-DA-PARTICIPACAO-EM-CONGRESSO/@TITULO",
414   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
415   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
416   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
417   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
418   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
419   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
420   "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
421 ]
422 },
423 {
424 "descricao" : "Participacao em seminarios",
425 "tag" : "/DADOS-COMPLEMENTARES/PARTICIPACAO-EM-EVENTOS-CONGRESSOS/
      PARTICIPACAO-EM-SEMINARIO",
426 "inicio" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-SEMINARIO/@ANO"],
427 "fim" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-SEMINARIO/@ANO"],
428 "atributos" : [
429   "/DADOS-BASICOS-DA-PARTICIPACAO-EM-SEMINARIO/@TITULO",
430   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
431   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
432   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
433   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
434   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
435   "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
436   "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
437 ]
438 },
439 {
440 "descricao" : "Participacao em simposios",
441 "tag" : "/DADOS-COMPLEMENTARES/PARTICIPACAO-EM-EVENTOS-CONGRESSOS/
      PARTICIPACAO-EM-SIMPOSIO",
442 "inicio" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-SIMPOSIO/@ANO"],
443 "fim" : ["/DADOS-BASICOS-DA-PARTICIPACAO-EM-SIMPOSIO/@ANO"],
444 "atributos" : [
```

```

445     "./DADOS-BASICOS-DA-PARTICIPACAO-EM-SIMPOSIO/@TITULO",
446     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
447     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
448     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
449     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
450     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
451     "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
452     "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
453 ]
454 },
455 {
456     "descricao" : "Participacao em encontros",
457     "tag" : "./DADOS-COMPLEMENTARES/PARTICIPACAO-EM-EVENTOS-CONGRESSOS/
PARTICIPACAO-EM-ENCONTRO",
458     "inicio" : ["./DADOS-BASICOS-DA-PARTICIPACAO-EM-ENCONTRO/@ANO"],
459     "fim" : ["./DADOS-BASICOS-DA-PARTICIPACAO-EM-ENCONTRO/@ANO"],
460     "atributos" : [
461         "./DADOS-BASICOS-DA-PARTICIPACAO-EM-ENCONTRO/@TITULO",
462         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
463         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
464         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
465         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
466         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
467         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
468         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
469     ]
470 },
471 {
472     "descricao" : "Orientacoes em andamento - Mestrado",
473     "tag" : "./DADOS-COMPLEMENTARES/ORIENTACOES-EM-ANDAMENTO/ORIENTACAO-
EM-ANDAMENTO-DE-MESTRADO",
474     "inicio" : ["./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-MESTRADO/
@ANO"],
475     "fim" : ["./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-MESTRADO/
@ANO"],
476     "atributos" : [
477         "./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-MESTRADO/@TITULO-
DO-TRABALHO",
478         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
479         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
480         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
481         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
482         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
483         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
484         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
485     ]
486 },
487 {
488     "descricao" : "Orientacoes em andamento - Doutorado",
489     "tag" : "./DADOS-COMPLEMENTARES/ORIENTACOES-EM-ANDAMENTO/ORIENTACAO-
EM-ANDAMENTO-DE-DOCTORADO",
490     "inicio" : ["./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-DOCTORADO/
@ANO"],
491     "fim" : ["./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-DOCTORADO/
@ANO"],
492     "atributos" : [
493         "./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-DOCTORADO/@TITULO-
DO-TRABALHO",
494         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
495         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
496         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
497         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
498         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
499         "./PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
500         "./INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
501     ]
502 },
503 {
504     "descricao" : "Orientacoes em andamento - Pos-Doutorado",
505     "tag" : "./DADOS-COMPLEMENTARES/ORIENTACOES-EM-ANDAMENTO/ORIENTACAO-
EM-ANDAMENTO-DE-POS-DOCTORADO",
506     "inicio" : ["./DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-POS-

```

```
    DOUTORADO/@ANO"],
507 "fim" : ["/DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-POS-
    DOUTORADO/@ANO"],
508 "atributos" : [
509     "/DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-POS-DOUTORADO/
        @TITULO-DO-TRABALHO",
510     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
511     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
512     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
513     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
514     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
515     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
516     "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
517 ]
518 },
519 {
520 "descricao" : "Orientacoes em andamento - Iniciacao Cientifica",
521 "tag" : "/DADOS-COMPLEMENTARES/ORIENTACOES-EM-ANDAMENTO/ORIENTACAO-
    EM-ANDAMENTO-DE-INICIACAO-CIENTIFICA",
522 "inicio" : ["/DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-INICIACAO-
    CIENTIFICA/@ANO"],
523 "fim" : ["/DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-INICIACAO-
    CIENTIFICA/@ANO"],
524 "atributos" : [
525     "/DADOS-BASICOS-DA-ORIENTACAO-EM-ANDAMENTO-DE-INICIACAO-
        CIENTIFICA/@TITULO-DO-TRABALHO",
526     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-1",
527     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-2",
528     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-3",
529     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-4",
530     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-5",
531     "/PALAVRAS-CHAVE/@PALAVRA-CHAVE-6",
532     "/INFORMACOES-ADICIONAIS/@DESCRICAO-INFORMACOES-ADICIONAIS"
533 ]
534 }
535 ]
```

APÊNDICE B – Trecho de um currículo Lattes em XML

Trecho referente ao XML do currículo <<http://lattes.cnpq.br/0055080134892269>>
(a representação ... indica partes do documento que foram suprimidas):

```

1 <CURRICULO-VITAE>
2 ...
3 <PRODUCAO-BIBLIOGRAFICA>
4 ...
5 <ARTIGOS-PUBLICADOS>
6 ...
7 <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="1364" ORDEM-IMPORTANCIA="22">
8 <DADOS-BASICOS-DO-ARTIGO
9 NATUREZA="COMPLETO"
10 TITULO-DO-ARTIGO="TRATAMENTO DE UM EFLUENTE MODELO TEXTIL VIA
    REACAO DE FENTON"
11 ANO-DO-ARTIGO="2016"
12 PAIS-DE-PUBLICACAO=""
13 IDIOMA="Portugues"
14 MEIO-DE-DIVULGACAO="MEIO_DIGITAL"
15 HOME-PAGE-DO-TRABALHO="[doi:10.15628/holos.2016.2250]"
16 FLAG-RELEVANCIA="NAO"
17 DOI="10.15628/holos.2016.2250"
18 TITULO-DO-ARTIGO-INGLES="TRATAMENTO DE UM EFLUENTE MODELO TEXTIL
    VIA REACAO DE FENTON"
19 FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
20
21 <DETALHAMENTO-DO-ARTIGO
22 TITULO-DO-PERIODICO-OU-REVISTA="Holos (Natal. Online)"
23 ISSN="18071600"
24 VOLUME="4"
25 FASCICULO=""
26 SERIE=""
27 PAGINA-INICIAL="25"
28 PAGINA-FINAL="35"
29 LOCAL-DE-PUBLICACAO=""/>
30 <AUTORES ... />
31 <AUTORES ... />
32 <AUTORES ... />
33 <PALAVRAS-CHAVE
34 PALAVRA-CHAVE-1="reacao fenton"
35 PALAVRA-CHAVE-2="efluente textil"
36 PALAVRA-CHAVE-3=""
37 PALAVRA-CHAVE-4=""
38 PALAVRA-CHAVE-5=""
39 PALAVRA-CHAVE-6="" />
40 <AREAS-DO-CONHECIMENTO>
41 <AREA-DO-CONHECIMENTO-1
42 NOME-GRANDE-AREA-DO-CONHECIMENTO="ENGENHARIAS"
43 NOME-DA-AREA-DO-CONHECIMENTO=""
44 NOME-DA-SUB-AREA-DO-CONHECIMENTO=""
45 NOME-DA-ESPECIALIDADE=""/>
46 </AREAS-DO-CONHECIMENTO>
47 <SETORES-DE-ATIVIDADE
48 SETOR-DE-ATIVIDADE-1="Pesquisa e desenvolvimento cientifico"
49 SETOR-DE-ATIVIDADE-2=""
50 SETOR-DE-ATIVIDADE-3=""/>
51 <INFORMACOES-ADICIONAIS
52 DESCRICAO-INFORMACOES-ADICIONAIS="O presente estudo teve como
    objetivo avaliar a aplicacao de um Processo Oxidativo Avancado
    no tratamento de efluentes texteis,denominado processo oxidativo
    de Fenton ..."/>
53 </ARTIGO-PUBLICADO>

```

```
54
55     ...
56
57     <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="786" ORDEM-IMPORTANCIA="104">
58     <DADOS-BASICOS-DO-ARTIGO
59         NATUREZA="COMPLETO"
60         TITULO-DO-ARTIGO="O ESTADO DA ARTE NA RECICLAGEM DE LAMPADAS
61             FLUORESCENTES NO BRASIL: PARTE 1"
62         ANO-DO-ARTIGO="2008"
63         PAIS-DE-PUBLICACAO=""
64         IDIOMA="Portugues"
65         MEIO-DE-DIVULGACAO="IMPRESSO"
66         HOME-PAGE-DO-TRABALHO=""
67         FLAG-RELEVANCIA="NAO"
68         DOI=""
69         TITULO-DO-ARTIGO-INGLES="O ESTADO DA ARTE NA RECICLAGEM DE LAMPADAS
70             FLUORESCENTES NO BRASIL: PARTE 1"
71         FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
72     <DETALHAMENTO-DO-ARTIGO
73         TITULO-DO-PERIODICO-OU-REVISTA="Acta Ambiental Catarinense"
74         ISSN="16777298"
75         VOLUME="5"
76         FASCICULO=""
77         SERIE="1"
78         PAGINA-INICIAL="43"
79         PAGINA-FINAL="53"
80         LOCAL-DE-PUBLICACAO="" />
81     <AUTORES ... />
82     <AUTORES ... />
83     <AUTORES ... />
84     <PALAVRAS-CHAVE
85         PALAVRA-CHAVE-1="lampadas fluoescntes"
86         PALAVRA-CHAVE-2="Reciclagem"
87         PALAVRA-CHAVE-3=""
88         PALAVRA-CHAVE-4=""
89         PALAVRA-CHAVE-5=""
90         PALAVRA-CHAVE-6="" />
91     <AREAS-DO-CONHECIMENTO>
92     <AREA-DO-CONHECIMENTO-1
93         NOME-GRANDE-AREA-DO-CONHECIMENTO="ENGENHARIAS"
94         NOME-DA-AREA-DO-CONHECIMENTO=""
95         NOME-DA-SUB-AREA-DO-CONHECIMENTO=""
96         NOME-DA-ESPECIALIDADE="" />
97     </AREAS-DO-CONHECIMENTO>
98     <SETORES-DE-ATIVIDADE
99         SETOR-DE-ATIVIDADE-1="Industrias de Transformacao"
100        SETOR-DE-ATIVIDADE-2=""
101        SETOR-DE-ATIVIDADE-3="" />
102     <INFORMACOES-ADICIONAIS
103         DESCRICAO-INFORMACOES-ADICIONAIS = "As lampadas fluoescntes sao
104             um artigo comum na sociedade atual. No entanto sao agentes de
105             contaminacao ambiental quando descartadas inadequadamente.
106             Algumas empresas tem se dedicado ao desenvolvimento de
107             tecnologias de descontaminacao deste material. ..."/>
108     </ARTIGO-PUBLICADO>
109     ...
110 </ARTIGOS-PUBLICADOS>
111 ...
112 </PRODUCAO-BIBLIOGRAFICA>
113 ...
114 </CURRICULO-VITAE>
```

APÊNDICE C – Stopwords da língua portuguesa

Abaixo está a lista de *stopwords* da língua portuguesa, que foi utilizada para limpar os documentos gerados a partir dos currículos Lattes. Ao todo, são 621 palavras compiladas com a junção de diferentes fontes de dados¹.

| | | | | |
|---------|----------|------------|-----------|-----------|
| a | aonde | breve | da | destes |
| acerca | aos | cada | daquela | deve |
| adeus | apenas | caminho | daquelas | devem |
| agora | apoio | catorze | daquele | devendo |
| ainda | apontar | cedo | daqueles | dever |
| alem | apos | cento | dar | deveria |
| algmas | após | certamente | das | deveriam |
| algo | aquela | certeza | de | deverá |
| algum | aquelas | cima | de | deverão |
| alguma | aquele | cinco | debaixo | devia |
| algumas | aqueles | coisa | dela | deviam |
| alguns | aqui | coisas | delas | dez |
| alguém | aquilo | com | dele | dezanove |
| ali | as | como | deles | dezasseis |
| além | assim | comprido | demais | dezassete |
| ambas | através | conhecido | dentro | dezoito |
| ambos | atrás | conselho | depois | dia |
| ampla | até | contra | desde | diante |
| amplas | aí | contudo | desligado | direita |
| amplo | baixo | corrente | dessa | dispoe |
| amplos | bastante | cuja | dessas | dispoem |
| ano | bem | cujas | desse | disse |
| anos | boa | cujo | desses | disso |
| ante | boas | cujos | desta | disto |
| antes | bom | custa | destas | dito |
| ao | bons | cá | deste | diversa |

¹ Foi utilizada a lista disponível na função `nlk.corpus.stopwords.words("portuguese")` e as listas dos sites <<https://gist.github.com/alopes/5358189>>, <<https://github.com/stopwords-iso/stopwords-pt/blob/master/stopwords-pt.txt>>, <<https://www.ranks.nl/stopwords/portuguese>>, <<https://virtuati.com.br/cliente/knowledgebase/25/Lista-de-StopWords.html>>

| | | | | |
|--------------|--------------|----------|-------------|----------|
| diversas | estavam | feita | houvera | maiorias |
| diversos | este | feitas | houveram | mais |
| diz | esteja | feito | houverei | mal |
| dizem | estejam | feitos | houverem | mas |
| dizer | estejamos | fez | houveremos | me |
| do | estes | fim | houveria | mediante |
| dois | esteve | final | houveriam | meio |
| dos | estive | foi | houvermos | menor |
| doze | estivemos | fomos | houverá | menos |
| duas | estiver | for | houverão | meses |
| durante | estivera | fora | houveríamos | mesma |
| dá | estiveram | foram | houvesse | mesmas |
| dão | estiverem | forem | houvessem | mesmo |
| dúvida | estivermos | forma | houvéramos | mesmos |
| e | estivesse | formos | houvéssemos | meu |
| ela | estivessem | fosse | há | meus |
| elas | estiveste | fossem | hão | mil |
| ele | estivestes | foste | iniciar | minha |
| eles | estivéramos | fostes | inicio | minhas |
| em | estivéssemos | fui | ir | momento |
| embora | estou | fôramos | irá | muita |
| enquanto | está | fôssemos | isso | muitas |
| entao | estás | geral | ista | muito |
| entre | estávamos | grande | iste | muitos |
| então | estão | grandes | isto | máximo |
| era | eu | grupo | já | mês |
| eram | exemplo | ha | la | na |
| essa | falta | haja | lado | nada |
| essas | fará | hajam | lhe | nao |
| esse | favor | hajamos | lhes | naquela |
| esses | faz | havemos | ligado | naquelas |
| esta | fazeis | havia | lo | naquele |
| estado | fazem | hei | local | naqueles |
| estamos | fazemos | hoje | logo | nas |
| estar | fazendo | hora | longe | nem |
| estar estará | fazer | horas | lugar | nenhum |
| estará | fazes | houve | lá | nenhuma |
| estas | fazia | houvemos | maior | nessa |
| estava | faço | houver | maioria | nessas |

| | | | | |
|----------|--------------|---------------|----------|----------|
| nesse | os | porquê | quarta | serão |
| nesses | ou | portanto | quarto | seríamos |
| nesta | outra | porém | quatro | sete |
| nestas | outras | posição | que | seu |
| neste | outro | possivelmente | quem | seus |
| nestes | outros | posso | quer | sexta |
| ninguém | para | possível | quereis | sexto |
| no | parece | pouca | querem | si |
| noite | parte | poucas | queremas | sido |
| nome | partir | pouco | queres | sim |
| nos | paucas | poucos | quero | sistema |
| nossa | pegar | povo | questão | sob |
| nossas | pela | primeira | quieto | sobre |
| nosso | pelas | primeiras | quinta | sois |
| nossos | pelo | primeiro | quinto | somente |
| nova | pelos | primeiros | quinze | somos |
| novas | pequena | projeto | quâis | sou |
| nove | pequenas | promeiro | quê | sua |
| novο | pequeno | propios | relação | suas |
| novos | pequenos | proprio | sabe | são |
| num | per | própria | sabem | sétima |
| numa | perante | próprias | saber | sétimo |
| numas | perto | próprio | se | só |
| nunca | peessoas | próprios | segunda | tal |
| nuns | pode | próxima | segundo | talvez |
| não | podem | próximas | sei | tambem |
| nível | podendo | próximo | seis | também |
| nós | poder | próximos | seja | tampouco |
| número | poderia | pude | sejam | tanta |
| o | poderiam | puderam | sejamos | tantas |
| obra | poderá | pôde | sem | tanto |
| obrigada | poderá podia | põe | sempre | tarde |
| obrigado | podia | põem | sendo | te |
| oitava | podiam | quais | ser | tem |
| oitavo | pois | qual | serei | temos |
| oito | ponto | qualquer | seremos | tempo |
| onde | pontos | quando | seria | tendes |
| ontem | por | quanto | seriam | tendo |
| onze | porque | quantos | será | tenha |

| | | | | |
|----------|------------|-----------|------------|---------|
| tenham | ti | todo | vai | vosso |
| tenhamos | tido | todos | vais | vossos |
| tenho | tinha | trabalhar | valor | vários |
| tens | tinham | trabalho | veja | vão |
| tentar | tipo | treze | vem | vêm |
| tentaram | tive | três | vendo | vós |
| tente | tivemos | tu | vens | zero |
| tentei | tiver | tua | ver | à |
| ter | tivera | tuas | verdade | às |
| terceira | tiveram | tudo | verdadeiro | área |
| terceiro | tiverem | tão | vez | é |
| terei | tivermos | tém | vezes | éramos |
| teremos | tivesse | têm | viagem | és |
| teria | tivessem | tínhamos | vindo | última |
| teriam | tiveste | ufsc | vinte | últimas |
| terá | tivestes | um | vir | último |
| terão | tivéramos | uma | você | últimos |
| teríamos | tivéssemos | umas | vocês | |
| teu | toda | uns | vos | |
| teus | todas | usa | vossa | |
| teve | todavia | usar | vossas | |

APÊNDICE D – Artigo

EXPER.TE: Um framework para identificação temporal da expertise

Luiz Eduardo Pizzinatto¹, Carina Friedrich Dorneles¹, Rodrigo Gonçalves¹

¹Departamento de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)
– Florianópolis – SC – Brasil

{luiz.pizzinatto, carina.dorneles, rodrigo.g}@ufsc.br,

Abstract. *Problems of knowledge discovery and information retrieval on the web are challenges that still exist, even with various already developed techniques. This paper deals specifically with expertise retrieval process. To achieve this goal, a framework is proposed: with unsupervised learning, the main researchers' production topics (registered on Lattes platform) are presented, taking into account trends and changes that have occurred over time. Each words set (from each topic) is later used to infer each researchers expertise. Temporal analysis is added to determine work relevance in relation to the period studied and how researcher's expertise evolves over time. In the end, the framework is tested with a small data set, with individual and joint analysis for all members of a Graduate Program.*

Resumo. *Problemas de descoberta do conhecimento e busca da informação na web são desafios que ainda existem, mesmo com as diversas técnicas já desenvolvidas. Este trabalho trata especificamente do processo de determinação da expertise. Como forma de facilitar esse processo, é apresentado um framework capaz de determinar, com aprendizado não-supervisionado, os principais tópicos das produções de pesquisadores cadastrados na plataforma Lattes, levando em consideração tendências e mudanças que ocorreram ao longo do tempo. As palavras que compõem os tópicos encontrados são, posteriormente, usadas para inferir a especialidade de cada indivíduo. A análise temporal é adicionada com o objetivo de determinar a relevância dos trabalhos em relação ao período pesquisado e como a expertise do pesquisador evolui ao longo do tempo. Ao final, o framework é testado com um pequeno conjunto de dados, com análise individual e em conjunto para todos os membros de um Programa de Pós Graduação.*

1. Introdução

A busca pela informação está fortemente presente no cotidiano das pessoas, desde suas ações mais básicas. A simples necessidade de buscar um número de telefone, encontrar uma receita ou mesmo uma foto, seja em meio físico ou digital, requer o acesso a uma base de dados para que o item desejado seja recuperado. Nas últimas décadas, dada a crescente massa de informação disponibilizada pelas Tecnologias de Informação e Comunicação, a Ciência da Computação e Ciência da Informação vêm abordando e discutindo esse assunto para encontrar meios de facilitar o acesso aos dados [Alves et al. 2007]. Dentro das necessidades de busca, existe uma específica que visa encontrar especialistas de uma

determinada área ou assunto. Pessoas têm procurado por especialistas desde antes do advento dos computadores e estes podem ser demandados não somente para responder a questionamentos de sua área, mas também para serem designados em alguma função ou trabalho, inclusive dentro de organizações [Balog et al. 2012]. Isso é fácil de ser exemplificado na área médica: os médicos são profissionais especialistas na grande área da saúde, mas é comum que possuam uma sub-especialidade, como pediatria, neurologia, dermatologia, cardiologia, sendo o mais adequado escolhido conforme o tipo de problema a ser resolvido. Da mesma maneira, na área da tecnologia, existem especialistas nas áreas de segurança da informação, infraestrutura de redes, bancos de dados, microcontroladores, entre várias outras. Encontrar o expert na área desejada é um dos grandes desafios da IR.

Considerando que a *web* é extremamente dinâmica, ignorar o fator *tempo* não é uma suposição realista a ser feita, pois a maioria dos conjuntos de dados (como artigos de pesquisadores, blogs e notícias) não possuem co-ocorrência estática; muito pelo contrário, padrões nos dados coletados estão sempre mudando, mostrando tendências de crescimento ou diminuição ao longo do tempo [Daud et al. 2009]. Dizer que uma pessoa *Y* é especialista no tópico *X* em 2010 pode não ser verdade em 2018.

Os meios acadêmico e corporativo quase sempre estão interessados em pessoas que possam ser consideradas especialistas em um determinado assunto no *momento atual*. No caso do meio acadêmico, pode ser interessante encontrar um professor ou pesquisador para: compor bancas de defesa de mestrado ou doutorado, criar questões ou compor bancas de concursos públicos, fazer parte de grupos de revisores de trabalhos submetidos em periódicos ou eventos científicos, compor uma nova linha de pesquisa dentro de um programa de pós-graduação, entre outros.

2. Conceitos

Nas próximas subseções, estão descritos os dados da fonte de dados utilizada e o modelo que integra o núcleo do framework.

2.1. Plataforma Lattes

A Plataforma Lattes representa a experiência do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na integração de bases de dados de Currículos, de Grupos de pesquisa e de Instituições em um único Sistema de Informações [CNPq 2018b]. Dentro dessa plataforma, o módulo específico para tratar dos currículos é conhecido como *Currículo Lattes*, formado por um sistema web onde os estudantes e pesquisadores registram sua vida pregressa de maneira padronizada, sendo utilizado pela maioria das instituições de fomento, universidades e institutos de pesquisa em suas ações de planejamento, gestão e operacionalização de fomento do CNPq. O [CNPq 2018a] disponibiliza periodicamente um arquivo com todos os IDs dos pesquisadores cadastrados no Lattes. No momento da extração para este trabalho, a base de dados era datada de 08/07/2017, contando com 5.197.727 currículos cadastrados, sendo que cada um pode ser exportado e baixado em formato XML. Este formato, além de conter todas as informações cadastradas (diferente da versão HTML que mostra parcialmente), as marcações conferem uma semântica muito mais precisa aos dados. A formatação deste XML segue uma estrutura definida por um Document Type Definition (DTD), disponibilizada pelo próprio CNPq.

2.2. Modelagem por tópicos e o TOT

A modelagem por tópicos é uma das técnicas utilizadas para extrair tópicos de maneira automática a partir de dados. Alguns esforços têm sido feitos nessa área utilizando métodos de clusterização, com objetivo de separar os documentos em grupos específicos, baseados em seu conteúdo semântico. Modelos probabilísticos de tópicos são um conjunto de algoritmos cujo objetivo é descobrir estruturas temáticas ocultas em grandes coleções de documentos [FALEIROS and LOPES 2016]. Quando várias palavras aparecem juntas com frequência (em documentos semanticamente relacionados), esse grupo pode ser considerado como um tópico. O Latent Dirichlet Allocation (LDA) é um desses modelos, sendo considerado um dos principais.

Os autores [Wang and McCallum 2006] apresentam um modelo de tópicos baseado no LDA, um dos principais algoritmos, chamado de *Topics over Time* (TOT), que é capaz de capturar não apenas as mudanças de baixa dimensão na estrutura dos dados, mas também como elas mudam ao longo do tempo. Neste modelo, os tópicos e os seus significados são modelados como constantes ao longo do tempo. O TOT captura mudanças na ocorrência (e coocorrência condicionada no tempo) dos próprios tópicos, não mudanças nas distribuições das palavras de cada tópico. Vale lembrar que da mesma forma que o LDA, o TOT considera os documentos como *bag-of-words*, portanto a ordem das palavras pode ser negligenciada, sendo capturada apenas sua coocorrência. Essa abordagem possui outras características, como ser capaz de dizer o ano do documento (de maneira muito aproximada) baseado em suas palavras. Na Figura 1 pode ser conferida representação visual do TOT, bem como a notação e parametrização utilizada na Tabela 1.

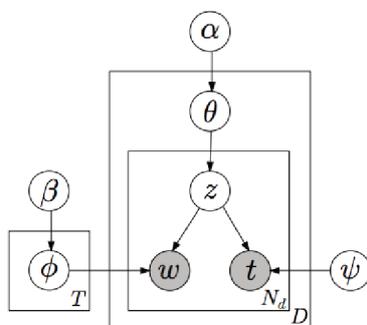


Figura 1. Representação visual (no formato *plate notation*) do TOT, com o amostrador de Gibbs. Fonte: [Wang and McCallum 2006]

3. O framework Exper.te

Para cumprir com os objetivos do trabalho, o framework Exper.te foi projetado para receber como entrada um arquivo XML de um currículo, fazer o processamento das informações e, ao final, fornecer dados que sejam suficientes para determinação da especialidade de um pesquisador ao longo do tempo. O encadeamento dos passos do framework pode ser conferido na Figura 2. Entre cada um dos passos, é indicado o tipo de informação que transita (saída do passo anterior e entrada para o próximo).

O primeiro passo a ser realizado é a leitura dos arquivos correspondentes aos currículos e processamento no formato XML (Passo 1.a), bem como extração dos da-

| Símbolo | Descrição |
|------------|--|
| T | número de tópicos |
| D | número de documentos |
| V | número de palavras únicas (considerando todos os documentos) |
| N_d | quantidade total de palavras no documento d |
| θ_d | distribuição multinomial de tópicos, específica do documento d |
| ϕ_z | distribuição multinomial de palavras, específica do tópico z |
| ψ_z | distribuição beta do tempo, específica do tópico z |
| z_{di} | tópico associado com a i^{a} palavra do documento d |
| t_{di} | o <i>timestamp</i> associado com a i^{a} palavra do documento d |

Tabela 1. Notação utilizada pelo TOT. Fonte: [Wang and McCallum 2006]

dos (Passo 1.b), que consistem basicamente na leitura dos arquivos texto, análise da estrutura e transformação em objetos XML (permitindo navegação e acesso aos dados). Em seguida, inicia-se a transformação do objeto XML, em um formato que possa ser utilizado pelo framework, gerando vários documentos (Passo 2) a partir da extração de informações relevantes. O XML é composto por vários elementos que correspondem à produção do pesquisador, incluindo artigos, projetos, orientações, atuações profissionais, formação acadêmica, entre outros. Cada ocorrência de elemento que corresponde a uma produção do pesquisador, com seus respectivos atributos, é usada na geração de documentos para o framework.

Como os dados do Currículo Lattes podem estar em idiomas diferentes do Português, é necessário fazer a tradução dos registros para um único idioma, com objetivo de normalizar os dados (Passo 3). A presença de diferentes idiomas ocorre, normalmente, pelo fato das produções estarem representadas no idioma do veículo no qual foram publicadas e, em menor parte, por ter o cadastro de pesquisadores de outros países. Como a plataforma Lattes possui quase sua totalidade de currículos referentes a pesquisadores brasileiros, a maior parte das informações estão registradas em Português. Por este motivo, optou-se pela tradução de todas as informações para o Português, com o intuito de diminuir a perda de informação e minimizar o esforço.

O Passo 4 é referente às limpezas realizadas no texto dos documentos já gerados. As entidades HTML são substituídas pelos seus caracteres visuais são removidas todas as pontuações e outros símbolos (Passo 4.a). Todas as letras são convertidas para minúscula, assim as comparações e alterações ficam padronizadas (Passo 4.b). O Passo 4.c faz identificação de palavras compostas, ou *collocations*, ou ainda bigramas, que são aquelas que ocorrem juntas com frequência (como é o caso de *banco dados*, *raio x*, *sistema operacional*). São palavras que fazem mais sentido quando estão juntas. No Passo 4.d é feita a remoção de palavras desnecessárias (ou palavras de ligação), também conhecidas na literatura por *stopwords*. Após a limpeza das palavras desnecessárias globais, que são fixas, é realizada uma estatística da quantidade de palavras para cada currículo, onde o usuário pode selecionar entre as mais recorrentes, quais serão removidas dos seus respectivos currículos (Passo 4.e). Essa etapa adicional se faz necessária pois podem aparecer palavras específicas do domínio de atuação da pessoa que não possuem grande

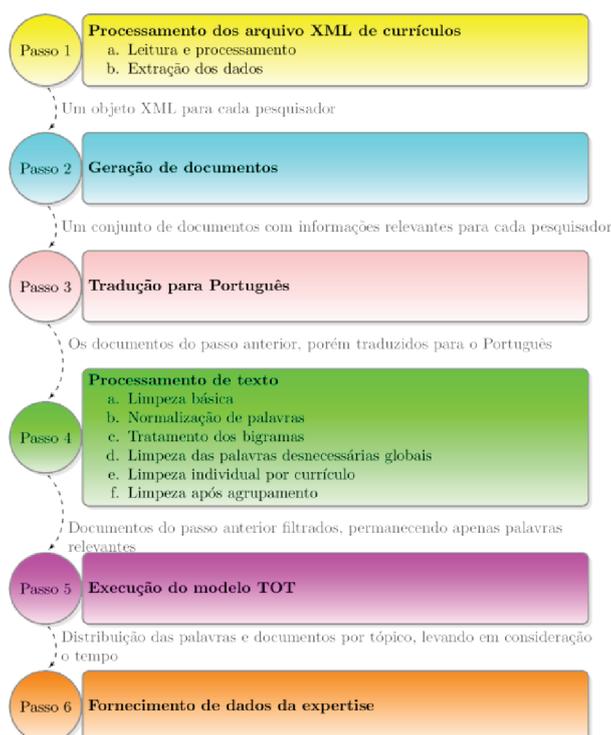


Figura 2. Representação do fluxo do framework. Fonte: Elaborado pelo autor.

significância. Nos casos de professores, alguns exemplos são *projeto*, *ensino*, *pesquisa*, *extensão*, entre outras. Quando vários currículos são analisados em conjunto, algumas palavras que não possuem força contexto no individual, acabam ganhando força maior no conjunto, mas ainda assim não possuem significado relevante. Neste caso, é feita uma nova rodada de contagem de palavras e eliminação das mais recorrentes (escolhidas pelo usuário), mas agora considerando todos os documentos (Passo 4.f).

O modelo de referência selecionado para este trabalho, que faz a recuperação da expertise considerando o tempo, foi o Topics Over Time (TOT). Após a transformação dos documentos, tradução e processamento do texto, é este modelo que receberá os documentos compilados para análise da expertise (Passo 5). Sua implementação foi realizada através de duas referências, uma do LDA em Java [Nguyen 2018] e outra, *opensource*, em Python, do próprio TOT [Maurya 2015]. Os gráficos, que auxiliam na interpretação dos resultados do TOT, foram gerados com a biblioteca *matplotlib* (Passo 6).

4. Experimentos e resultados

Para cumprir com os objetivos do trabalho, é definida uma metodologia para execução dos experimentos, que envolve a seleção de Programas de Pós Graduação, escolha de pesquisadores para análise individual e definição dos parâmetros de execução.

Para validar o framework com os currículos de maneira individual, foram selecionados 3 currículos de 3 Programas de Pós Graduação da UFSC, sendo que neste trabalho será relatado apenas um caso. Para analisar um grupo de pesquisadores, foi escolhido o Programa de Pós Graduação em Ciência da Computação (PPGCC), pela familiaridade do autor com os membros do programa. Neste sentido, o trabalho considerou todos aqueles que constavam no Sistema de Controle Acadêmico da Pós-Graduação (CAPG) como

membros do programa, até agosto de 2018, sendo coletados 27 currículos. Como a característica do conjunto dados é diferente da análise individual (com mais dados e mais campos de pesquisa envolvidos), os parâmetros de execução foram alterados, conforme pode ser verificado na Tabela 2. O valor de α mais alto se deve ao fato de que, sabidamente, existem mais tópicos no conjunto de currículos analisados (comparado a uma análise individual), portanto, os documentos provavelmente possuem uma mistura maior dos tópicos existentes. Como existem mais palavras no conjunto analisado, considerou-se um valor menor de β , para que os tópicos tenham uma mistura menor das palavras disponíveis. O número de tópicos foi considerado como $T = 10$, uma vez que existem 7 linhas de pesquisa dentro do programa, o valor foi levemente aumentado para comportar possíveis variações dentro do algoritmo. Valores menores que 7 não fazem sentido, escondendo tópicos ou forçando uma relação irreal na composição. Valores muito altos sofrem do mesmo problema relatado no caso das análises individuais (palavras ficam esparsas e tópicos sem sentido são criados). Valor igual a sete também não permite que tópicos um pouco mais fracos apareçam (por exemplo: dois tópicos fortes de Banco de Dados podem tomar o lugar de um tópico de Computação Paralela e Distribuída). Durante os experimentos, o valor 10 se mostrou mais adequado para o conjunto de dados.

O período de início e fim foi considerado o mesmo da criação do programa até o momento do início desta pesquisa. Mesmo que os atuais integrantes provavelmente não sejam os mesmos de 1992, é importante verificar se eles estiveram e continuam aderentes às linhas de pesquisa. Todos os valores foram aproximados a partir de testes realizados com os dados, que foram considerados aceitáveis para o conjunto de dados.

As linhas de pesquisa estão divididas do PPGCC estão divididos da seguinte maneira: 1. Computação Paralela e Distribuída (a. Computação em Nuvem; b. Computação de Alto Desempenho; c. Web Services; d. Tolerância a Falhas; e. Sistemas de Tempo Real; f. Internet das Coisas; g. Paralelização de Algoritmos); 2. Banco de Dados (a. Big Data; b. Mineração de Dados; c. Data Science; d. Recuperação de Informação; e. Gerenciamento de Dados Complexos; f. Processamento de Linguagem Natural; g. Web Semântica), 3. Engenharia de Software (a. Gerência de Projetos; b. Computação na Escola; c. Métodos Ágeis; d. Modelagem de Dados; e. Métodos Formais Aplicados; f. Interação Humana-Computador), 4. Inteligência Computacional (a. Computação Gráfica; b. Reconhecimento de Padrões; c. Aprendizado de Máquina; d. Sistemas Multiagentes; e. Robótica; f. Agentes Inteligentes; g. Inteligência Artificial), 5. Redes de Computadores (a. Sistemas Distribuídos; b. Protocolos de Comunicação; c. Redes de Sensores; d. Internet das Coisas; e. Computação em Nuvem), 6. Segurança em Sistemas Computacionais (a. Segurança Cibernética; b. Segurança em Sistemas Distribuídos; c. Criptografia; d. Certificação Digital; e. Cripto-moedas; f. Blockchain; g. Privacidade), 7. Sistemas Embarcados (a. Redes de Sensores; b. Internet das Coisas; c. Microeletrônica; d. Processamento de Sinais Digitais; e. Automação de Projeto Eletrônico).

4.1. Análise

No caso do currículo individual, da área de Ciência da Computação, foram gerados $D = 736$ documentos, sendo que a etapa de tradução detectou 6 idiomas (Português, Inglês, Francês, Espanhol, Galego e Hmong). Para ele, a divisão dos 6 tópicos e as palavras com maior probabilidade de ocorrer em cada um são exibidas no mapa de calor da Figura 3. Percebe-se a correlação forte de todos os tópicos com a palavra *rede*, principal-

| Parâmetro | Valor inicial (PPGCC) | Valor inicial (individual) |
|----------------------------------|-----------------------|----------------------------|
| α | 0.15 | 0.10 |
| β | 0.03 | 0.10 |
| ψ_z (tópico x tempo) | $(a, b) = (1, 1)$ | $(a, b) = (1, 1)$ |
| θ_d (documento x tópico) | 0 | 0 |
| ϕ_z (palavra x tópico) | 0 | 0 |
| Iterações do amostrador de Gibbs | 1000 | 1000 |
| T (quantidade de tópicos) | 10 | 6 |
| Ano inicial | 1992 | 0 |
| Ano final | 2018 | 2018 |

Tabela 2. Parâmetros utilizados na inicialização do TOT na execução para os currículos do PPGCC e individual. Fonte: Elaborado pelo autor.

mente envolvendo gerência. As relações passam por computação em grade nos tópicos tópicos 1 e 3, ambos com viés de segurança; gerência de redes (tópicos 3, 5 e 6), sendo o tópico 5 com viés de sustentabilidade em nuvens-verdes; nuvem (tópicos 2 e 5); gerência com viés de monitoramento e qualidade de serviço ligadas a redes sem fio ad hoc (tópico 4). É visível que *segurança* (relacionada também com *detecção-intrusão*) permeia vários tópicos, muito provável pela questão da *gerência* estar fortemente presente (que indica, entre outras coisas, cuidado com a segurança). Na página do Currículo Lattes, existe o seguinte trecho informado pelo próprio autor: *Tem experiência na área de Ciência da Computação e Telecomunicações, com ênfase em Administração e Gerência de Redes e Serviços, atuando principalmente nos seguintes temas: segurança, computação autônoma, computação em nuvem e internet das coisas.*

Sobre a temporalidade, os tópicos 4 e 6 estão levemente deslocados no tempo. Avaliando a composição de palavras, um parece a evolução do outro, ou, minimamente, uma variação. Os tópicos mais proeminentes são o 2 e 5, que envolvem justamente o conceito de *cloud computing*, um com mais enfoque em segurança e outro em sustentabilidade. O gráfico de probabilidades acumuladas da Figura 5 mostra que os tópicos 1, 3, 4 e 6 coexistiram por algum tempo, mas os esforços nos últimos anos estão voltados para os tópicos 2 e 5.

Para a execução com todos os currículos do PPGCC, as principais palavras de cada tópico podem ser conferidas na Figura 6. Para fazer uma análise com esses tópicos, foram utilizadas as linhas de pesquisa do Programa e suas subdivisões, sendo destacado os seguintes pontos:

- Sobre o tópico 1, pode-se dizer que está relacionado exclusivamente com a linha 3, tratando de processos de software, que envolvem planejamento, melhoria, qualidade, sendo o CMMI um exemplo de processo, tendo a ISO-IEC definido vários outros processos;
- Sobre o tópico 2, há uma semelhança muito grande com os elementos da linha 2, principalmente com o 2.f - Processamento de Linguagem Natural e 2.g - Web semântica, passando também por outros elementos da mesma linha;
- No tópico 3, há uma aderência com a linha de pesquisa 4, passando pelo 4.b -

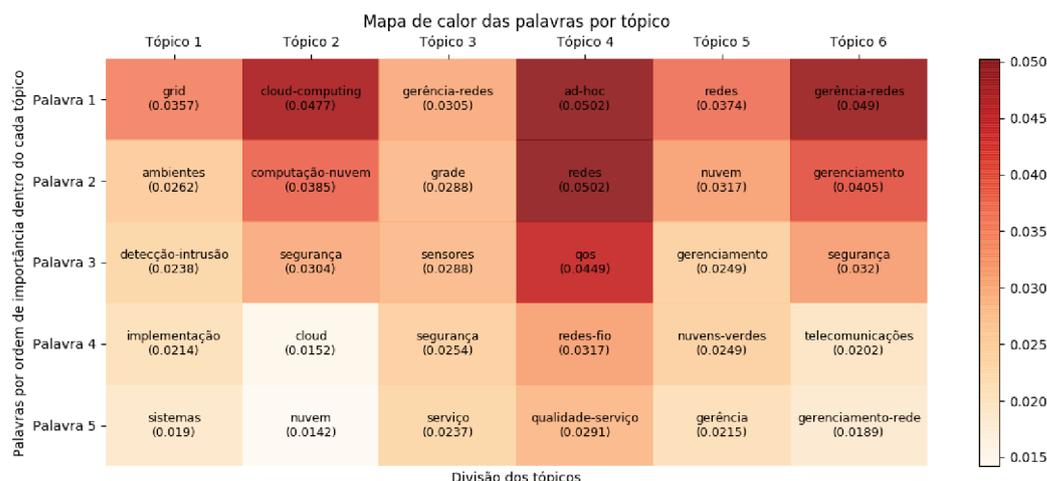


Figura 3. Mapa de calor das palavras por tópico do currículo individual. Fonte: Elaborado pelo autor.

Reconhecimento de Padrões (que envolve o 4.c - Aprendizado de Máquina), 4.d - Sistemas Multi Agentes e 4.f - Agentes Inteligentes, que são todos itens da 4.g - Inteligência Artificial;

- Com relação ao tópico 5, parece haver relação com a linha 3, com mais proximidade ao 3.b - Computação na escola e com 3.f - Interação Humana-Computador;
- No tópico 6, há uma forte relação com a linha de pesquisa 4, envolvendo os itens 4.1, 4.b, 4.c e 4.d (e inerentemente o 4.g). Os assuntos desse tópico (imagens e telemedicina) são abordados por pesquisadores do departamento de ensino onde a maioria dos professores do PPGCC estão lotados;
- O tópico 7 parece ter sido isolado para tratar do item 3.b, com nuances do. Aqui aparecem algumas palavras que sugerem o assunto do tópico 6. Isso provavelmente deve ter acontecido pela proximidade dos pesquisadores que tratam dos dois assuntos;
- O tópico 8 tem forte relação com a linha 6, mais especificamente com 6.a e 6.d, também com a linha 5, e uma mistura da linha 1, que podem ser considerados correlacionados;
- Os dois últimos tópicos são os que apresentam as misturas, tornando-os mais abrangentes e menos específicos. No caso do tópico 9, há presença das linhas de pesquisa 2, 3, 4 e 5 (misturando banco de dados, com redes, inteligência artificial, web e engenharia de software), e no tópico 10 a presença das linhas de pesquisa 3 (design, computação na escola, modelagem), 4 (sistemas multiagentes) e 7 (microeletrônica, timing, circuitos integrados, systemc).

Pela análise acima, é possível perceber que todas as linhas de pesquisa foram contempladas de alguma forma, com destaque para a área 3 - Engenharia de Software e 4 - Inteligência Computacional, que permeiam vários tópicos. Vale lembrar que elas podem ter sido potencializadas pelas demais áreas, uma vez que são ortogonais. Inteligência Artificial, por exemplo, pode ser utilizada em redes de computadores para deteção de intrusão, em banco de dados para data mining, em sistemas embarcados para melhoria de

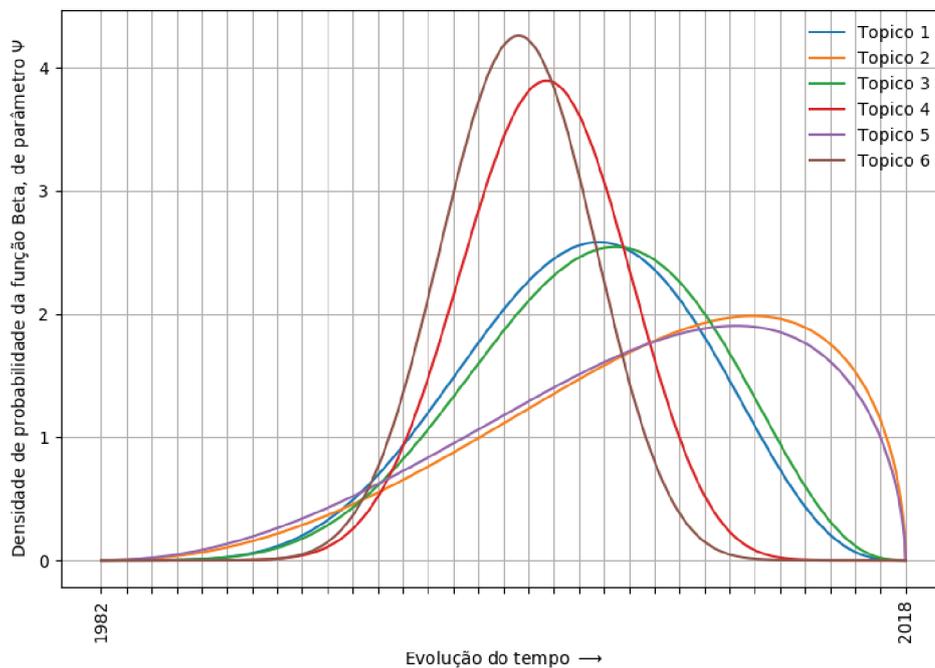


Figura 4. Evolução dos tópicos ao longo do tempo do currículo individual. Fonte: Elaborado pelo autor.

projetos digitais, na nuvem para detecção de falhas, entre outros. Engenharia de Software pode estar presente na modelagem de dados, melhoria de processos, análise de requisitos, gestão, usabilidade, verificação, validação e testes de qualquer uma das outras linhas de pesquisa (inclusive de sistemas embarcados).

Analisando a temporalidade nos gráficos das Figuras 7 e 8, inicialmente se destacam justamente os dois tópicos que apresentam maior mistura de linhas de pesquisa (9 e 10). Eles aparecem isolados, aproximadamente na metade da vida do programa. A mistura deve ter acontecido por algum padrão de coocorrência das palavras nessa época, que não se repetiu nos anos seguintes. Os demais tópicos parecem seguir a mesma tendência ao longo do tempo, com produção relativamente constante entre eles. Um destaque é para o assunto do tópico 7, que parece ter surgido no início do 4º quartil do espaço de tempo, tendo produções cada vez maiores, assim como uma diminuição gradativa da produção no tópico 6. Por outro lado, o tópico 3 (que parece tratar de reconhecimento de padrões, assim como o 6), está em alta nos últimos anos. Neste caso, pode ser adicionada a interpretação de que está ocorrendo uma mudança no foco desse tópico, onde os assuntos das publicações estão sendo alteradas gradativamente.

5. Trabalhos futuros

O processamento do texto de entrada é peça fundamental para geração de bons resultados. Seria interessante realizar um estudo mais aprofundado de técnicas de limpeza de dados, assim os dados teriam mais relevância. Um passo de processamento de texto que não foi incorporado a este trabalho é a normalização morfológica: na Língua Portuguesa, as palavras podem variar de acordo com gênero (masculino ou feminino), número (singular

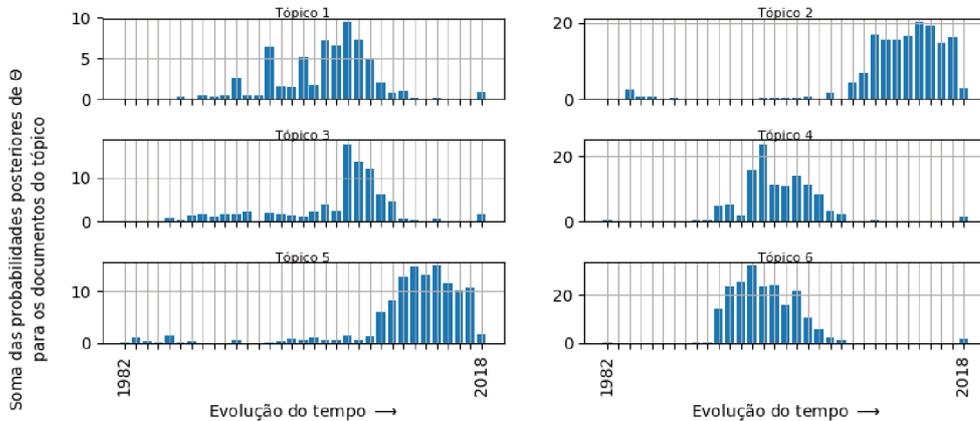


Figura 5. Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, do currículo individual. Fonte: Elaborado pelo autor.

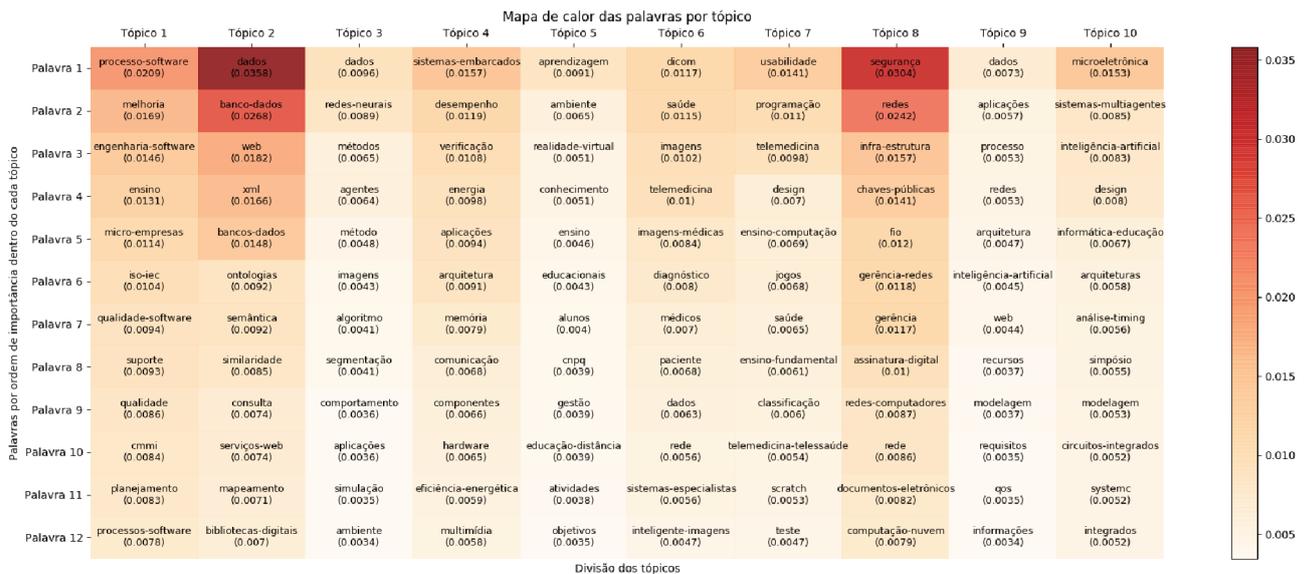


Figura 6. Mapa de calor das palavras por tópico de todos os currículos do PPGCC. Fonte: Elaborado pelo autor.

ou plural), grau (aumentativo ou diminutivo), pessoa (primeira, segunda ou terceira pessoa do singular ou plural), modo (indicativo, subjuntivo e imperativo), tempo (presente, pretérito e futuro), voz (ativa, passiva e reflexiva), conforme a classificação gramatical, e ainda possuir prefixos ou sufixos. Essa característica faz com que palavras como *rede* e *redes* sejam diferentes, tendo menos importância individualmente. Caso elas fossem transformadas para sua raiz antes da análise, ambas poderiam ser tratadas apenas como *rede*.

A busca por especialistas que possuam similaridade entre si ou com um grupo também pode ser implementada utilizando a função de divergência de Kullback–Leibler

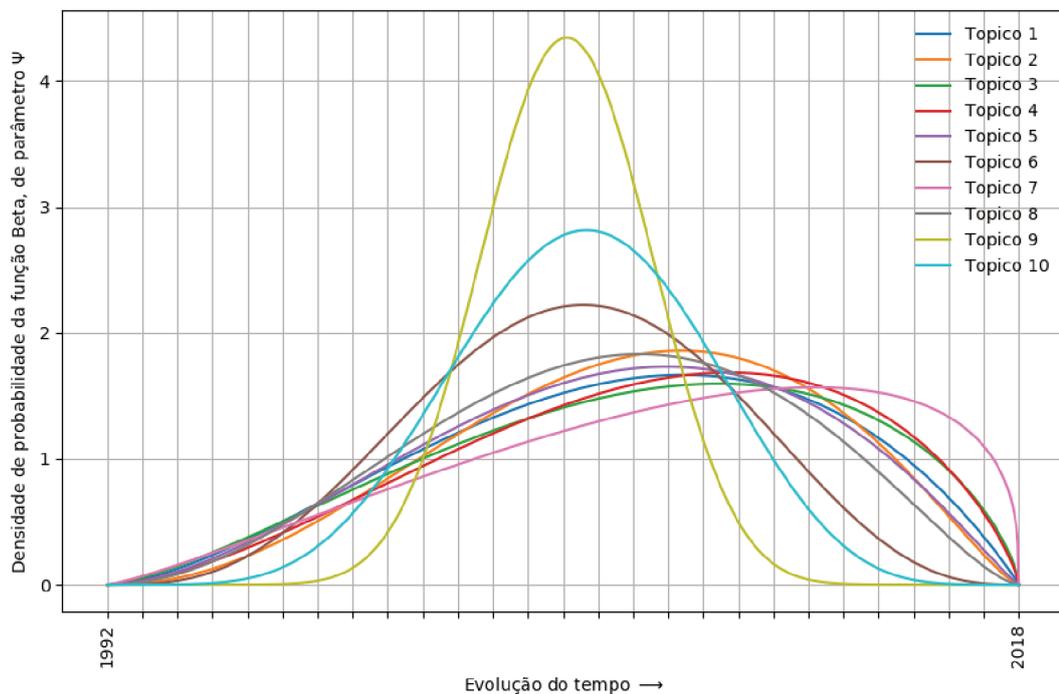


Figura 7. Evolução dos tópicos ao longo do tempo de todos os currículos do PPGCC. Fonte: Elaborado pelo autor.

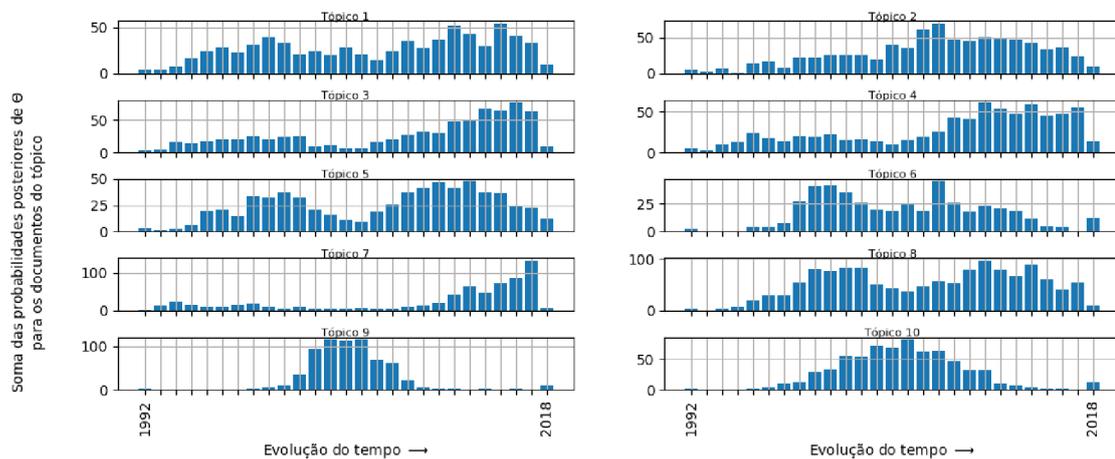


Figura 8. Gráfico da probabilidade posterior acumulada θ por tópico ao longo do tempo, de todos os currículos do PPGCC. Fonte: Elaborado pelo autor.

(ou divergência KL). Por exemplo, ao aplicar o TOT para um pesquisador qualquer, sua similaridade pode ser comparada ao PPGCC através da divergência KL, para saber a proximidade deste pesquisador com as linhas do programa. A métrica KL pode ser aplicada entre resultados da execução do TOT.

Outro ponto que pode ser evoluído é o modelo utilizado. Como discutido nas seções iniciais, o TOT possui uma abordagem de tópicos imutáveis. Isso confere uma certa rigidez e sensibilidade aos tópicos gerados (alterações pequenas geram novos tópicos). Existem modelos mais robustos, como o TAT, que são capazes de capturar também a mutação dos tópicos, quer permite que algumas palavras ganhem força em detrimento de outras ao longo do tempo, dentro do mesmo tópico.

Caso um modelo baseado em LDA continue sendo utilizado, pode ser realizado um estudo aprofundado nas técnicas para determinação do número ideal de tópicos, bem como parametrização adequada de α e β .

6. Conclusão

Para verificação do framework, foram feitos testes com 3 pesquisadores diferentes, e posterior aplicação no simultaneamente em 27 pesquisadores do PPGCC. Neste ponto fica evidente a influência do tempo, onde alguns tópicos emergem em detrimento de outros. Mais especificamente no caso do Programa de Pós Graduação, os tópicos de maior relação com as linhas de pesquisa permanecem mais constantes no tempo, como era esperado, pois precisam se manter (mesmo que um de seus integrantes mude o foco da pesquisa, a linha permanece com demais membros pesquisando na mesma área).

Duas das dificuldades foram (i) encontrar o valor ideal de tópicos e dos parâmetros α e β e (ii) efetuar um filtro correto nas palavras. Com relação ao filtro, a ação está condicionada na granularidade desejada por quem está analisando os dados: quanto mais específica a necessidade dos dados, mais as palavras genéricas precisam ser removidas dos documentos. No caso dos parâmetros, foram encontrados baseados em testes consecutivos, pela familiaridade com os dados analisados, ajustando-os pouco a pouco. Nada impede que os valores genéricos propostos na literatura sejam usados, mas um ajuste fino pode trazer uma grande melhora.

Por fim, a ferramenta apresentada mostra-se como uma opção para análise da temporalidade em conjuntos de dados, com uma possível inclusão nos processos de gestão de pesquisadores. Para que o modelo seja mais robusto e possa atender de forma efetiva e real, os itens propostos nos trabalhos futuros devem ser considerados e tratados.

Referências

- Alves, R. C. V., Banhos, V. T. M., Bicheri, A. L. A., Segundo, J. E. S., and Woida, L. M. (2007). Ciência da informação, ciência da computação e recuperação da informação: algumas considerações sobre os métodos e tecnologias da informação utilizados ao longo do tempo. *Revista Eletrônica Informação e Cognição*, 6(1):28–40.
- Balog, K., Fang, Y., and de Rijke, M. (2012). *Expertise Retrieval*. Foundations and Trends in Information Retrieval. Now Publishers.
- CNPq (2018a). Lattes - extração de dados.
- CNPq (2018b). A plataforma lattes.
- Daud, A., Li, J., Zhou, L., and Muhammad, F. (2009). Exploiting temporal authors interests via temporal-author-topic modeling. In Huang, R., Yang, Q., Pei, J., Gama, J., Meng, X., and Li, X., editors, *Advanced Data Mining and Applications*, pages 435–443, Berlin, Heidelberg. Springer Berlin Heidelberg.
- FALEIROS, T. D. P. and LOPES, A. D. A. (2016). Modelos probabilísticos de tópicos: desvendando o latent dirichlet allocation. Technical Report 409, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.
- Maurya, A. (2015). Topics over time - an open-source implementation.
- Nguyen, D. Q. (2018). jLDADMM: A Java package for the LDA and DMM topic models. *arXiv preprint arXiv:1808.03835*.
- Wang, X. and McCallum, A. (2006). Topics over time: A non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 424–433, New York, NY, USA. ACM.

APÊNDICE E – Fontes

Arquivo <main.py>.

```

1 import argparse
2
3 from temporal_analysis import TemporalAnalysis
4
5 parser = argparse.ArgumentParser(description='Executa o framework que
6     identifica a expertise dos pesquisadores',
7     formatter_class=argparse.ArgumentDefaultsHelpFormatter)
8
9 parser.add_argument('-f', '--file',
10     required=True,
11     nargs='+',
12     help="""Arquivo XML do Currículo Lattes a ser processado.
13     Junto com cada arquivo será gerado outro de mesmo nome, com
14     extensão .dat,
15     com os dados utilizados gerados para o currículo em específico.
16     Caso o arquivo .dat exista, ele será considerado. Para
17     desconsiderar os arquivos .dat
18     prévios, utilizar a opção -r """)
19
20 parser.add_argument('-r', '--regenerate',
21     action="count",
22     default=0,
23     help='Desconsidera os arquivos .dat de cada currículo, gerando
24     novamente.');
```

```

25 parser.add_argument('-tc', '--translate_credentials',
26     help="""Arquivo contendo as credenciais usadas pela API do Google
27     para tradução.
28     Mais informações em https://cloud.google.com/docs/authentication/production#auth-cloud-implicit-python""");
29
30 parser.add_argument('-l', '--lattes_structure',
31     default="./lattes_dtd.json",
32     help='Arquivo JSON contendo a estrutura de TAGs do Lattes a ser
33     considerada');
```

```

34 parser.add_argument('-a', '--alpha',
35     default=0.1,
36     type=float,
37     help='Valor inicial do parâmetro alpha. Baseado na parametrização
38     do LDA.');
```

```

39 parser.add_argument('-b', '--beta',
40     default=0.1,
41     type=float,
42     help='Valor do parâmetro beta. Baseado na parametrização do LDA.')
```

```

43 parser.add_argument('-i', '--iterations',
44     default=400,
45     type=int,
46     help='Quantidade de iterações a serem executadas do Amostrador de
47     Gibbs');
```

```

48 parser.add_argument('-sw', '--stop_words',
49     required=True,
50     nargs='+',
51     help='Arquivos contendo as stopwords padrões, que serão usadas em
52     todos os currículos. Arquivos texto, contendo uma palavra por
53     linha');
```

```

54 parser.add_argument('-nw', '--number_words_show',
55     default=100,
56     type=int,
```

```
53     help='Numero de palavras a serem exibidas para seleáo durante o
54         filtro de cada curráculo. Neste momento serão selecionadas as
55         stopwords individuais para cada curráculo. Valor zero indica
56         que não haverá filtro.');
```

```
54 parser.add_argument('-t', '--topics',
55                     default=6,
56                     type=int,
57                     help='Quantidade de táticos a serem gerados');
```

```
58 parser.add_argument('-s', '--start_year',
59                     default=0,
60                     type=int,
61                     help='Ano inicial (inclusive) para considerar os documentos (
62                         documentos de anos anteriores serão desconsiderados).
63                         Influencia apenas no momento da execuáo do TOT.');
```

```
64
65 parser.add_argument('-e', '--end_year',
66                     default=0,
67                     type=int,
68                     help='Ano final (inclusive) para considerar os documentos (
69                         documentos de anos posteriores serão desconsiderados).
70                         Influencia apenas no momento da execuáo do TOT.');
```

```
71 parser.add_argument('-o', '--output_file',
72                     required=True,
73                     help='Arquivo com o resultado da execuáo do algoritmo');
```

```
74
75 parser.add_argument('-v', '--verbose',
76                     action="count",
77                     default=0,
78                     help='Ativa o log da execuáo no terminal.');
```

```
79
80 args = parser.parse_args()
81
82 analysis = TemporalAnalysis(
83     args.file,
84     args.regenerate,
85     args.lattes_structure,
86     args.alpha,
87     args.beta,
88     args.iterations,
89     args.topics,
90     args.stop_words,
91     args.number_words_show,
92     args.translate_credentials,
93     args.start_year,
94     args.end_year,
95     args.output_file,
96     args.verbose)
97
98 analysis.run()
```

Script com exemplo de uso.

```
1 #!/bin/sh
2
3 python3 main.py -f \
4   ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
5   /2828124381670231_1.xml" \
6   ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
7   /7644756660823271_1.xml" \
8   ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
9   /0378897709136226_1.xml" \
10  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
11  /1235242182279350_1.xml" \
12  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
13  /4386203755404108_1.xml" \
14  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
15  /3879944876244096_1.xml" \
16  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
17  /5048342535708224_1.xml" \
18  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
19  /6246567808516505_1.xml" \
20  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
21  /3531795815749652_1.xml" \
22  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
23  /3431795837830476_1.xml" \
24  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
25  /2020489905881170_1.xml" \
26  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
27  /5029700010533047_1.xml" \
28  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
29  /7115792628126490_1.xml" \
30  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
31  /6876016315737507_1.xml" \
32  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
33  /5872119613051645_1.xml" \
34  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
35  /8496058927371924_1.xml" \
36  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
37  /2212288622038633_1.xml" \
38  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
39  /0911364223406030_1.xml" \
40  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
41  /7541399131195077_1.xml" \
42  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
43  /6405951782839858_1.xml" \
44  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
45  /6609568015287399_1.xml" \
46  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
47  /8018788492846096_1.xml" \
48  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
49  /9716092379282146_1.xml" \
50  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
51  /1154078371182103_1.xml" \
52  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
53  /5011370918857999_1.xml" \
54  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
55  /7879026373930318_1.xml" \
56  ../expertise_retrieval/crawlers/baixaLattes/Import_20180527_1638
57  /9045199822095908_1.xml" \
58  -tc ./TccTranslation-0f3e6b735223.json \
59  -t 10 \
60  -a 0.15 \
61  -b 0.03 \
62  -l lattes_dtd.json \
63  -i 1000 \
64  -s 1992 \
65  -e 2018 \
66  -sw ./stopwords01.txt \
67  ./stopwords02.txt \
68  ./stopwords03.txt \
69  ./stopwords04.txt \
70  ./stopwords05.txt \
```

```
44 -o out.txt \  
45 -vvv
```

Arquivo <main_plot.py>.

```
1 import argparse
2 import pickle
3
4 from tot import TopicsOverTimeVisualization
5
6 parser = argparse.ArgumentParser(description='Faz a plotagem dos dados
7     de uma determinada execução do framework',
8     formatter_class=argparse.ArgumentDefaultsHelpFormatter)
9 parser.add_argument('-f', '--file',
10     required=True,
11     help='Arquivo que armazenou o resultado da execução do TOT');
12
13 args = parser.parse_args()
14
15 par = pickle.load(open(args.file, "rb" ))
16
17 TopicsOverTimeVisualization(par)
```

Arquivo <temporal_analysis.py>.

```
1 import json
2
3 from lattes import LattesTOT
4 from logger import Logger
5 from text_process import BasicClean, ProfileClean, ProfileCollocations,
   GeneralClean
6 from translate import Translator
7 from xml_load import XMLLoad
8
9
10 class TemporalAnalysis:
11     def __init__(self,
12                 files, regenerate, lattes_structure,
13                 alpha, beta, iterations, topics,
14                 stop_words, number_words_show,
15                 translate_credentials,
16                 start_year,
17                 end_year,
18                 output_file,
19                 verbose):
20         self.files = files
21         self.regenerate = regenerate
22         self.lattes_structure = lattes_structure
23         self.alpha = alpha
24         self.beta = beta
25         self.iterations = iterations
26         self.topics = topics
27         self.stop_words = stop_words
28         self.number_words_show = number_words_show
29         self.translate_credentials = translate_credentials
30         self.start_year = start_year
31         self.end_year = end_year
32         self.output_file = output_file
33         self.log = Logger(verbose)
34
35
36
37     def run(self):
38         self.log("Iniciando")
39
40         self.log("Carregando DTD do Lattes")
41         self.load_dtd()
42
43         self.log("Iniciando parser dos arquivos XML")
44         docs = XMLLoad(self.dtd, self.files, self.log).load()
45
46         self.log("Iniciando traducao dos documentos")
47
48         Translator(docs, self.translate_credentials, log = self.log).translate
49             ()
50
51         self.log("Remoção da pontuação e stopwords básicas")
52         BasicClean(docs, self.stop_words, self.log).clean()
53
54         self.log("Cálculo e substituição dos bigramas mais recorrentes")
55         ProfileCollocations(docs, self.log).replace_bigrams()
56
57         self.log("Fazendo limpeza específica por currículo")
58         ProfileClean(docs, self.number_words_show, self.log).clean()
59
60         self.log("Fazendo limpeza específica de todos juntos")
61         GeneralClean(docs, self.number_words_show, self.log).clean()
62
63         self.log("Iniciando o tot")
64         LattesTOT(docs, self.alpha, self.beta, self.iterations, self.topics,
65                 self.start_year, self.end_year, self.output_file, self.log).run()
66
67
68
```

```
69 def load_dtd(self):  
70     f = open(self.lattes_structure)  
71     self.dtd = json.load(f)
```

Arquivo <lattes.py>.

```

1 import pickle
2
3 from logger import Logger
4 from tot import TopicsOverTime, TopicsOverTimeVisualization
5
6 class LattesTOT:
7
8     def __init__(self, documents, alpha, beta, iterations, topics,
9                 start_year, end_year, output_file, log=Logger()):
10        self.documents = documents
11        self.alpha = alpha
12        self.beta = beta
13        self.iterations = iterations
14        self.topics = topics
15        self.start_year = start_year
16        self.end_year = end_year
17        self.output_file = output_file
18        self.log = log
19
20    def run(self):
21
22        docs = []
23
24        for file in self.documents:
25
26            for doc in file["docs"]:
27                if (int(doc["start"]) >= self.start_year) and (int(doc["start"]) <=
28                    self.end_year):
29                    docs.append({
30                        "authors": [file["name"]],
31                        "year": doc["start"],
32                        # "words": doc["doc_clean"]
33                        "words": doc["current"].split()
34                    });
35
36        documents = []
37        timestamps = []
38        dictionary = set()
39        sorted_docs = sorted(docs, key=lambda x: x["year"])
40
41        for doc in sorted_docs:
42            documents.append(doc["words"])
43            timestamps.append(int(doc["year"]))
44            dictionary.update(set(doc["words"]))
45
46        dictionary = list(dictionary)
47
48        #normalizando os timestamps
49        unique_timestamps = set(timestamps)
50        self.log("Unique timestamps: " + str(unique_timestamps), 5)
51        first_timestamp = timestamps[0]
52        last_timestamp = timestamps[len(timestamps)-1]
53        timestamps = [1.0*(t-first_timestamp)/(last_timestamp-first_timestamp)
54                    for t in timestamps]
55
56        self.log("Quantidade de documentos: " + str(len(documents)), 5)
57        self.log("Tamanho do dicionário: " + str(len(dictionary)), 5)
58
59        tot = TopicsOverTime(documents,
60                            timestamps,
61                            dictionary,
62                            self.alpha,
63                            self.beta,
64                            self.iterations,
65                            self.topics,
66                            self.log)
67
68        theta, phi, psi = tot.TopicsOverTimeGibbsSampling()
69        par = tot.parameters

```

```
69
70 topics_year_words = [{} for _ in range(par["T"])]
71 topics_year_posterior_probability = [{} for _ in range(par["T"])]
72
73 tot = 0
74
75 for t in range(par["T"]):
76     for d in range(len(documents)):
77
78         year = str(sorted_docs[d]["year"])
79
80         if year not in topics_year_words[t]:
81             topics_year_words[t][year] = 0
82             topics_year_posterior_probability[t][year] = 0.0
83
84         topics_year_words[t][year] += par['m_original'][d][t
85             ] #words
86         topics_year_posterior_probability[t][year] += par['m'][d][t] #prob
87
88 par["histogram_wordsxtopicxyear"] = topics_year_words
89 par["histogram_probxtopicxyear"] = topics_year_posterior_probability
90 par["unique_timestamps"] = unique_timestamps
91
92
93
94 tot_pickle = open(self.output_file, 'wb')
95 pickle.dump(par, tot_pickle)
96 tot_pickle.close()
97
98 TopicsOverTimeVisualization(par)
```

Arquivo <logger.py>.

```
1 class Logger:
2     def __init__(self, level = -1):
3         self.level = level
4
5     def __call__(self, str, level = 1):
6         if (level <= self.level):
7             print(str)
```

Arquivo <text_process.py>.

```

1 import nltk
2 import re
3 import statistics
4 import string
5
6
7 from logger import Logger
8 from nltk import FreqDist
9 from nltk.corpus import stopwords
10 from nltk.collocations import *
11 from selection_list import SelectionList
12
13
14 class BasicClean:
15
16     def __init__(self, documents, stop_words_files, log = Logger()):
17         self.documents = documents
18         self.exclude = set(string.punctuation)
19         self.log = log
20         self.compile_stop_words(stop_words_files)
21         self.log("Pontuacao a ser removida: " + str(self.exclude), 3)
22
23
24     def clean(self):
25         for file in self.documents:
26
27             for docs in file["docs"]:
28                 if "doc_clean" not in docs:
29                     doc = docs["current"]
30
31                     punc_free = ''.join(ch if ch not in self.exclude else " " for ch in
32                                     doc.lower())
33                     stop_free = " ".join([i for i in punc_free.split() if i not in self
34                                     .stop_words])
35                     number_free = re.sub(r"\S*\d+\S*", "", stop_free) #removendo as
36                                     palavras que contenham números
37
38                     docs["doc_clean"] = number_free
39                     docs["current"] = number_free
40
41                     self.log("Antes limpeza: " + docs["doc"], 5)
42                     self.log("Depois limpeza: " + docs["current"], 5)
43                 else:
44                     docs["current"] = docs["doc_clean"]
45
46
47     def compile_stop_words(self, stop_words_files):
48         words = set()
49         words.update(stopwords.words("portuguese"))
50
51         for f in stop_words_files:
52             file = open(f, "r")
53             l = file.read().splitlines()
54             words.update(l)
55             file.close()
56
57         self.log("Stop words compiladas " + str(words) , 3)
58
59         self.stop_words = words
60
61
62 class ProfileCollocations:
63     def __init__(self, documents, log = Logger()):
64         self.documents = documents
65         self.log = log
66
67     def threshold_data(self, dic):
68         data = []
69         for w in sorted(dic, key=dic.get, reverse=False):
70             data.append(float(dic[w]))
71         mean = statistics.mean(data)
72         stdev = statistics.pstdev(data)

```

```

69     return mean + 2*stdev
70
71     def replace_bigrams(self):
72         for file in self.documents:
73             self.log("Copiando doc_clean para doc_clean_collocation", 3)
74             full_text = ""
75
76             #juntando todos os documentos do mesmo curriculo
77             for doc in file["docs"]:
78                 full_text += doc["current"].strip() + " "
79
80             self.log("Calculando os bigramas do arquivo " + file["original_file"]
81                    ], 2);
82             bigram_measures = nltk.collocations.BigramAssocMeasures()
83             finder = BigramCollocationFinder.from_words(full_text.split())
84
85             self.log("Calculando as ocorrencias dos bigramas", 2)
86             scores = {}
87             for p,s in finder.score_ngrams(bigram_measures.pmi):
88                 if (len(p[0])+ len(p[1]) > 3 ):
89                     composite_word = " ".join(p)
90                     scores[composite_word] = full_text.count(composite_word)
91
92             self.log("Quantidade de bigramas detectados: " + str(len(scores)), 3)
93
94             #juntando as collocations com mais de TRESHOLD
95             #para evitar dupla substituição, isso á:
96             #coleta-organizaááo-publicaááo dado-produááo-cientáfica-tácnica
97             #docentes
98             #substituiu primeiro coleta-organizaááo e depois organizaááo-publicaá
99             #áo
100            #certifica-se que a substituição não inicia nem termina com háfen
101            #e dá preferência para os collocation que aparecem mais vezes em
102            #todos esses documentos juntos
103            # substitui apenas uma vez. Isso á, se vai gerar "coleta-organizaááo
104            #publicaááo" ou "coleta organizaááo-publicaááo"
105            #depende da ordem que aparecerá no vetor de frequencia
106            #para isso usa a regeg (?<!--)\bPALAVRA1 PALAVRA2\b(?!--)
107            #os \b indicam os limites de uma palavra
108            #o (?<!--) serve para indicar que não deve comeáar com - (lookahead)
109            #o (?!--) indica que não pode terminar com -
110            TRESHOLD = self.threshold_data(scores)
111            file["bigrams_threshold"] = TRESHOLD
112            file["bigrams"] = []
113            self.log("Substituindo os bigramas com mais de {} ocorrências.".
114                   format(TRESHOLD), 3)
115            for w in sorted(scores, key=scores.get, reverse=True):
116                if (scores[w] >= TRESHOLD):
117                    collocation = w.replace(" ", "-") # "junta" a palavra com hifen
118                    file["bigrams"].append((collocation, scores[w]))
119                    for doc in file["docs"]:
120                        #doc["doc_clean_collocation"] = doc["doc_clean_collocation"].
121                        #replace(w, collocation)
122                        doc["current"] = re.sub(r"(?<!--)\b" + w + r"\b(?!--)", collocation,
123                                               doc["current"])
124
125            else:
126                break
127
128            self.log("Unificando o doc_clean_collocation e eliminando palavras
129                    com len(palavra) < 3", 2)
130
131            for doc in file["docs"]:
132                doc["current"] = " ".join([w for w in doc["current"].split() if (len
133                    (w) >= 3)])
134                doc["doc_clean_collocation"] = doc["current"]
135
136            self.log("***** Após remoááo dos bigramas e palavras tamanho < 3
137                    *****", 3)
138            self.log(doc["current"], 3)
139            self.log("*****", 3)
140
141     class ProfileClean:

```

```
130 def __init__(self, documents, number_words_show, log = Logger()):
131     self.documents = documents
132     self.number_words_show = number_words_show
133     self.log = log
134
135 def clean(self):
136     for file in self.documents:
137         all_words = []
138         for d in file["docs"]:
139             d["current"] = d["current"].split()
140             all_words += d["current"]
141
142         frequency = FreqDist(all_words)
143
144         frequency.most_common(self.number_words_show)
145
146         self.log("Fazendo a limpeza para o arquivo {} ({}).format(file["
147             original_file"], file["name"]))
148
149         stopwords = []
150         if ("stopwords" in file):
151             stopwords = file["stopwords"]
152
153         file["stopwords"] = SelectionList(frequency.most_common(self.
154             number_words_show), stopwords).choose()
155
156         for d in file["docs"]:
157             d["current"] = " ".join([w for w in d["current"] if w not in file["
158                 stopwords"]])
159             d["words_doc_clean_without_stopwords"] = d["current"]
160             self.log("Após limpeza individual: " + str(d["
161                 words_doc_clean_without_stopwords"]), 2)
162
163         file.save()
164
165 class GeneralClean:
166 def __init__(self, documents, number_words_show, log = Logger()):
167     self.documents = documents
168     self.number_words_show = number_words_show
169     self.log = log
170
171 def clean(self):
172     all_words = []
173     for file in self.documents:
174         for d in file["docs"]:
175             d["current"] = d["current"].split()
176             all_words += d["current"]
177
178         frequency = FreqDist(all_words)
179
180         frequency.most_common(self.number_words_show)
181
182         stopwords = SelectionList(frequency.most_common(self.number_words_show
183             ), []).choose()
184
185         for file in self.documents:
186             for d in file["docs"]:
187                 d["current"] = " ".join([w for w in d["current"] if w not in
188                     stopwords])
189                 d["words_doc_clean_without_general_stop_words"] = d["current"]
190                 self.log("Após limpeza geral: " + str(d["
191                     words_doc_clean_without_general_stop_words"]), 2)
```

Arquivo <translate.py>.

```

1 from google.cloud import translate
2 from logger import Logger
3
4 class Translator:
5
6     def __init__(self, documents, translate_credentials,
7                 max_char_per_request = 4000, target_language = 'pt', log = Logger())
8         :
9         self.documents = documents
10        self.translate_credentials = translate_credentials
11        self.max_char_per_request = max_char_per_request
12        self.target_language = target_language
13        self.log = log
14
15    def translate(self):
16        for file in self.documents:
17            docs = file["docs"]
18
19            size_current_request = 0;
20
21            docs_to_translate = []
22
23            qt_docs = len(docs)
24            last_doc = qt_docs - 1
25
26            for d in range(qt_docs):
27                #não chequei o doc_translated individual pelo manuseio do indice (
28                #se tiver um documento no meio que nao foi traduzido). Refatorar
29                #isso.
30                if ("translated" not in file or not file["translated"]):
31                    doc = docs[d]["current"]
32                    docs_to_translate.append(doc)
33                    size_current_request += len(doc)
34
35                if (size_current_request >= self.max_char_per_request or d ==
36                    last_doc):
37                    qt_docs_to_translate = len(docs_to_translate)
38
39                    try:
40                        translate_client = translate.Client.from_service_account_json(
41                            self.translate_credentials)
42                        self.log("Tentando traduzir {} documentos com {} caracteres
43                            ({} / {})".format(qt_docs_to_translate, size_current_request, d
44                            +1, qt_docs), 2)
45                        translations = translate_client.translate(docs_to_translate,
46                            target_language=self.target_language)
47
48                        for i in range(len(translations)):
49                            t = translations[i]
50                            src = t["detectedSourceLanguage"]
51                            index_original = d - (qt_docs_to_translate - 1) + i
52                            docs[index_original]["doc_translated"] = t["translatedText"]
53                            docs[index_original]["current"] = t["translatedText"]
54                            docs[index_original]["src_idiom"] = src
55
56                    except Exception as e:
57                        self.log("Não conseguiu traduzir os {} documentos contendo {}
58                            caracteres".format(qt_docs_to_translate, size_current_request
59                            ), 2)
60                        self.log(e, -1);
61
62                    size_current_request = 0
63                    docs_to_translate = []
64                else:
65                    docs[d]["current"] = docs[d]["doc_translated"]
66
67            #depois de tudo, salva
68            file["translated"] = True
69            file.save()

```


Arquivo <xml_load.py>.

```

1 from logger import Logger
2 from lxml import etree as et
3
4 import datetime
5 import html
6 import os
7 import pickle
8 import re
9
10
11 class XMLLoad:
12     def __init__(self, dtd, files, log = Logger()):
13         self.dtd = dtd
14         self.files = files
15         self.log = log
16
17     def load(self):
18         data = [];
19         for f in self.files:
20             self.log("Tentando carregar o arquivo " + f)
21             dat_file = File(f)
22
23             if (not dat_file.loaded()):
24                 xml_file = et.parse(f)
25                 docs = self.compile_documents(xml_file)
26                 dat_file["docs"] = docs
27                 dat_file["name"] = xml_file.xpath("./DADOS-GERAIS/@NOME-COMPLETO")
28                     [0]
29
30             data.append(dat_file)
31
32     return data
33
34     def compile_documents(self, element):
35         current_year = str(datetime.datetime.now().year)
36
37         group = 0
38         docs = []
39
40         #para cada trecho de informacao necessária do dtd
41         for d in self.dtd:
42             self.log("\n *****" + d["descricao"] + "*****", 2)
43             #encontra todas as tags que casam com esse item
44             for e in element.findall(d["tag"]):
45                 group += 1 #cada "tag" é considerado um grupo de informaáées. Todos
46                 os documentos gerados com esse grupo levam o mesmo id de grupo
47                 start = self.find_first_matching_element(e, d["inicio"], "****")
48                 end = self.find_first_matching_element(e, d["fim"], current_year)
49
50                 doc = ""
51
52                 #todos os atributos dentro dessa tag vão compor o documento
53                 for attr_xpath in d["atributos"]:
54                     attr = e.xpath(attr_xpath)
55                     for a in attr:
56                         a = a.replace("&";", "&")
57                         a = html.unescape(a)
58                         a = a.replace("\n", " ").replace("\r", " ").strip()
59                         if (len(a)):
60                             if (len(doc)):
61                                 doc += ". " + a
62                             else:
63                                 doc = a
64
65                 if (len(doc)):
66                     full_doc = {
67                         "type": d["descricao"],
68                         "tag": d["tag"],
69                         "start": start,
70                         "end": end,
71                         "group": group,

```

```
70     "doc": doc,
71     "current": doc
72 }
73
74     self.log(full_doc, 4)
75
76     docs.append(full_doc)
77     return docs
78
79 def find_first_matching_element(self, element, paths, default = ""):
80     for periodo_xpath in paths:
81         xp = element.xpath(periodo_xpath)
82         if xp and len(xp) and len(xp[0]):
83             return xp[0]
84     return default
85
86
87 class File(dict):
88     def __init__(self, original_file):
89         filename, fileext = os.path.splitext(original_file)
90         dat = filename + ".dat"
91
92     try:
93         fileDat = open(dat, "rb")
94         data = pickle.load(fileDat)
95         fileDat.close()
96         super().__init__(data)
97         self["loaded"] = True
98     except:
99         self["loaded"] = False
100         super().__init__({
101             "original_file": original_file,
102             "dat_file": dat
103         })
104
105
106 def save(self):
107     fileDat = open(self["dat_file"], "wb")
108     pickle.dump(self, fileDat, pickle.HIGHEST_PROTOCOL)
109     fileDat.close()
110
111 def loaded(self):
112     return self["loaded"]
```

Arquivo <selection_list.py>.

```

1 class SelectionList:
2     def __init__(self, items, selected):
3         self.items = items
4         self.selected = set(selected)
5
6
7     def choose(self):
8         size_index = len(str(len(self.items)))
9         longest = max([len(i[0]) for i in self.items]) + 1
10
11         choosing = True
12
13         while choosing:
14             print("*** Escolha as opções abaixo ***")
15             index = 1
16             for i in self.items:
17                 str_index = str(index)
18                 str_word = str(i[0])
19                 while len(str_index) < size_index:
20                     str_index = " " + str_index
21
22                 while len(str_word) < longest:
23                     str_word += " "
24
25                 line = str_index
26
27                 if (i[0] in self.selected):
28                     line += " * - "
29                 else:
30                     line += "   - "
31
32                 line += str_word + "(" + str(i[1]) + ")"
33
34                 index += 1
35
36                 print(line)
37
38             print ("\nPalavras já selecionadas: {} \n".format(", ".join(self.
39                 selected)))
40             options = input("Informe o número das opções (separados por espaços,
41                 negativo para excluir, vazio para terminar)\nou\nInforme as
42                 palavras por extenso (separadas por espaço, negativo para excluir)
43                 : ")
44
45             if (options != ""):
46                 for o in options.split():
47                     try:
48                         o = int(o)
49                         index = abs(o) - 1
50
51                         if (o > 0):
52                             self.selected.add(self.items[index][0])
53                         else:
54                             self.selected.remove(self.items[index][0])
55                     except:
56                         if (o[:1] == "-"): #se começa com -
57                             self.selected.remove(o[-(len(o)-1):]) #índice negativo pega de tr
58                                 ás pra frente
59                         else:
60                             self.selected.add(o)
61
62                 else:
63                     choosing = False
64
65             return self.selected

```

Arquivo <tot.py>.

```

1 # Copyright 2015 Abhinav Maurya
2
3 # This program is free software: you can redistribute it and/or modify
4 # it under the terms of the GNU General Public License as published by
5 # the Free Software Foundation, either version 3 of the License, or
6 # (at your option) any later version.
7
8 # This program is distributed in the hope that it will be useful,
9 # but WITHOUT ANY WARRANTY; without even the implied warranty of
10 # MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
11 # GNU General Public License for more details.
12
13 # You should have received a copy of the GNU General Public License
14 # along with this program. If not, see <http://www.gnu.org/licenses/>.
15
16 # Adapted in 2018 - Luiz Eduardo Pizzinatto final course paper
17
18 import math
19 import matplotlib.pyplot as plt
20 import numpy as np
21 import random
22
23 import scipy.special
24 import scipy.stats
25
26
27 from copy import deepcopy
28 from logger import Logger
29 from math import log
30 from scipy.stats import beta
31
32 class TopicsOverTime:
33
34     def __init__(self, documents, timestamps, dictionary, alpha, beta,
35                 iterations, topics, logger = Logger()):
36         self.log = logger
37         par = {} # dictionary of all parameters
38         par['dataset'] = 'pnas' # dataset name
39         par['max_iterations'] = iterations # max number of iterations in gibbs
40         # sampling
41         par['T'] = topics # number of topics
42         par['D'] = len(documents)
43         par['V'] = len(dictionary)
44         par['N'] = [len(doc) for doc in documents]
45         par['alpha_original'] = alpha
46         par['beta_original'] = beta
47         par['alpha'] = [alpha for _ in range(par['T'])]
48         par['beta'] = [beta for _ in range(par['V'])]
49         par['beta_sum'] = sum(par['beta'])
50         par['psi'] = [[1 for _ in range(2)] for _ in range(par['T'])]
51         par['betafunc_psi'] = [scipy.special.beta(par['psi'][t][0], par['psi']
52         ][t][1]) for t in range(par['T'])]
53         par['word_id'] = {dictionary[i]: i for i in range(len(dictionary))}
54         par['word_token'] = dictionary
55         par['z'] = [[random.randrange(0, par['T']) for _ in range(par['N'][d])]
56         for d in range(par['D'])]
57         par['t'] = [[timestamps[d] for _ in range(par['N'][d])] for d in range
58         (par['D'])]
59         par['w'] = [[par['word_id'][documents[d][i]] for i in range(par['N'][d])
60         ] for d in range(par['D'])]
61         par['m'] = [[0 for t in range(par['T'])] for d in range(par['D'])]
62         par['n'] = [[0 for v in range(par['V'])] for t in range(par['T'])]
63         par['n_sum'] = [0 for t in range(par['T'])]
64         np.set_printoptions(threshold=np.inf)
65         np.seterr(divide='ignore', invalid='ignore')
66
67         self.parameters = par
68         self.CalculateCounts()
69
70     def CalculateCounts(self):
71         par = self.parameters

```

```

66     for d in range(par['D']):
67         for i in range(par['N'][d]):
68             topic_di = par['z'][d][i] #topic in doc d at position i
69             word_di = par['w'][d][i] #word ID in doc d at position i
70             par['m'][d][topic_di] += 1
71             par['n'][topic_di][word_di] += 1
72             par['n_sum'][topic_di] += 1
73
74
75     def GetTopicTimestamps(self):
76         par = self.parameters
77         topic_timestamps = []
78         for topic in range(par['T']):
79             current_topic_timestamps = []
80             current_topic_doc_timestamps = [[ (par['z'][d][i]==topic)*par['t'][d
81                 ][i] for i in range(par['N'][d])] for d in range(par['D'])]
82             for d in range(par['D']):
83                 current_topic_doc_timestamps[d] = filter(lambda x: x!=0,
84                     current_topic_doc_timestamps[d])
85             for timestamps in current_topic_doc_timestamps:
86                 current_topic_timestamps.extend(timestamps)
87             assert current_topic_timestamps != []
88             topic_timestamps.append(current_topic_timestamps)
89         return topic_timestamps
90
91     def GetMethodOfMomentsEstimatesForPsi(self):
92         par = self.parameters
93         topic_timestamps = self.GetTopicTimestamps()
94         psi = [[1 for _ in range(2)] for _ in range(len(topic_timestamps))]
95         for i in range(len(topic_timestamps)):
96             current_topic_timestamps = topic_timestamps[i]
97             timestamp_mean = np.mean(current_topic_timestamps)
98             timestamp_var = np.var(current_topic_timestamps)
99             if timestamp_var == 0:
100                 timestamp_var = 1e-6
101             common_factor = timestamp_mean*(1-timestamp_mean)/timestamp_var - 1
102             psi[i][0] = 1 + timestamp_mean*common_factor
103             psi[i][1] = 1 + (1-timestamp_mean)*common_factor
104         return psi
105
106     def ComputePosteriorEstimatesOfThetaAndPhi(self):
107         par = self.parameters
108         theta = deepcopy(par['m'])
109         phi = deepcopy(par['n'])
110
111         for d in range(par['D']):
112             if sum(theta[d]) == 0:
113                 theta[d] = np.asarray([1.0/len(theta[d]) for _ in range(len(theta[d]
114                     ]))])
115             else:
116                 theta[d] = np.asarray(theta[d])
117                 theta[d] = 1.0*theta[d]/sum(theta[d])
118         theta = np.asarray(theta)
119
120         for t in range(par['T']):
121             if sum(phi[t]) == 0:
122                 phi[t] = np.asarray([1.0/len(phi[t]) for _ in range(len(phi[t]))])
123             else:
124                 phi[t] = np.asarray(phi[t])
125                 phi[t] = 1.0*phi[t]/sum(phi[t])
126         phi = np.asarray(phi)
127
128         return theta, phi
129
130     def ComputePosteriorEstimatesOfTheta(self, par):
131         theta = deepcopy(par['m'])
132
133         for d in range(par['D']):
134             if sum(theta[d]) == 0:
135                 theta[d] = np.asarray([1.0/len(theta[d]) for _ in range(len(theta[d]
136                     ]))])
137             else:

```

```

134     theta[d] = np.asarray(theta[d])
135     theta[d] = 1.0*theta[d]/sum(theta[d])
136
137     return np.matrix(theta)
138
139 def ComputePosteriorEstimateOfPhi(self, par):
140     phi = deepcopy(par['n'])
141
142     for t in range(par['T']):
143         if sum(phi[t]) == 0:
144             phi[t] = np.asarray([1.0/len(phi[t]) for _ in range(len(phi[t]))])
145         else:
146             phi[t] = np.asarray(phi[t])
147             phi[t] = 1.0*phi[t]/sum(phi[t])
148
149     return np.matrix(phi)
150
151 def TopicsOverTimeGibbsSampling(self):
152     par = self.parameters
153     for iteration in range(par['max_iterations']):
154         for d in range(par['D']):
155             for i in range(par['N'][d]):
156                 word_di = par['w'][d][i]
157                 t_di = par['t'][d][i]
158
159                 old_topic = par['z'][d][i]
160                 par['m'][d][old_topic] -= 1
161                 par['n'][old_topic][word_di] -= 1
162                 par['n_sum'][old_topic] -= 1
163
164                 topic_probabilities = []
165                 for topic_di in range(par['T']):
166                     psi_di = par['psi'][topic_di]
167                     topic_probability = 1.0 * (par['m'][d][topic_di] + par['alpha'][
168                         topic_di])
169                     topic_probability *= ((1-t_di)**(psi_di[0]-1)) * ((t_di)**(psi_di
170                         [1]-1))
171                     topic_probability /= par['betafunc_psi'][topic_di]
172                     topic_probability *= (par['n'][topic_di][word_di] + par['beta'][
173                         word_di])
174                     topic_probability /= (par['n_sum'][topic_di] + par['beta_sum'])
175                 topic_probabilities.append(topic_probability)
176                 sum_topic_probabilities = sum(topic_probabilities)
177                 if sum_topic_probabilities == 0:
178                     topic_probabilities = [1.0/par['T'] for _ in range(par['T'])]
179                 else:
180                     topic_probabilities = [p/sum_topic_probabilities for p in
181                         topic_probabilities]
182
183                 new_topic = list(np.random.multinomial(1, topic_probabilities, size
184                     =1)[0]).index(1)
185                 par['z'][d][i] = new_topic
186                 par['m'][d][new_topic] += 1
187                 par['n'][new_topic][word_di] += 1
188                 par['n_sum'][new_topic] += 1
189
190                 if d%10 == 0:
191                     self.log('Done with iteration {iteration} and document {document}'.
192                         format(iteration=iteration, document=d))
193                 par['psi'] = self.GetMethodOfMomentsEstimatesForPsi()
194                 par['betafunc_psi'] = [scipy.special.beta(par['psi'][t][0], par['psi
195                     '][t][1]) for t in range(par['T'])]
196                 par['m_original'] = deepcopy(par['m'])
197                 par['n_original'] = deepcopy(par['n'])
198                 par['m'], par['n'] = self.ComputePosteriorEstimatesOfThetaAndPhi()
199                 return par['m'], par['n'], par['psi']
200
201 class TopicsOverTimeVisualization():

```

```

199 def __init__(self, par):
200     self.HeatMap(par['n'], par['word_token'], par['T'], 8)
201     #self.VisualizeTopics(par['n'], par['word_token'], par['T'])
202     self.VisualizeEvolution(par['psi'], par["unique_timestamps"])
203     #self.VisualizeHistogram(par["histogram_wordsxtopicxyear"], "
        Quantidade de palavras por tapico x palavra x tempo")
204     self.VisualizeHistogram(par["histogram_probxtopicxyear"], "Soma das
        probabilidades posteriores de " + r"$\Theta$" + "\npara os
        documentos do tapico")
205
206
207 def HeatMap(self, phi, words, num_topics, num_words_per_topic):
208     left_labels = ["Palavra {}".format(i+1) for i in range(
        num_words_per_topic)]
209     top_labels = []
210     values = []
211     texts = []
212     for topic_index in range(len(phi)):
213         top_labels.append("Tapico {}".format(topic_index + 1))
214         print("Tapico {} (sum: {})".format(topic_index + 1, sum(phi[
            topic_index])))
215
216         topic_values = []
217         topic_texts = []
218         for original_word_index, value in sorted(enumerate(phi[topic_index]),
            key=lambda x:x[1], reverse=True)[:num_words_per_topic]:
219             topic_values.append(phi[topic_index][original_word_index])
220             topic_texts.append("{}\n{}".format(words[original_word_index], int
                (phi[topic_index][original_word_index]*10000)/10000.0))
221             #topic_texts.append("{}".format(words[original_word_index]))
222
223         values.append(topic_values)
224         texts.append(topic_texts)
225
226     values = np.transpose(values)
227     texts = np.transpose(texts)
228
229
230     fig, ax = plt.subplots()
231     #im = ax.imshow(values)
232     heatmap = plt.pcolor(values, cmap=plt.cm.OrRd, alpha=0.8)
233     plt.colorbar()
234     plt.ylabel('Palavras por ordem de importancia dentro do cada tapico',
        multialignment='center')
235     plt.xlabel('Divisao dos tapicos', multialignment='center')
236
237
238     ax.grid(False)
239     ax.set_frame_on(False)
240
241     ax.invert_yaxis()
242     ax.xaxis.tick_top()
243
244     # We want to show all ticks...
245     ax.set_xticks(np.arange(len(top_labels)) + 0.5)
246     ax.set_yticks(np.arange(len(left_labels)) + 0.5)
247     # ... and label them with the respective list entries
248     ax.set_xticklabels(top_labels, minor=False)
249     ax.set_yticklabels(left_labels, minor=False)
250
251     # Loop over data dimensions and create text annotations.
252     for i in range(len(left_labels)):
253         for j in range(len(top_labels)):
254             text = ax.text(j + 0.5, i + 0.5, texts[i, j], ha="center", va="
                center", color="k", fontsize=9)
255
256     ax.set_title("Mapa de calor das palavras por tapico")
257     fig.tight_layout()
258     plt.show()
259
260
261

```

```

262 def VisualizeTopics(self, phi, words, num_topics, viz_threshold=2e-2):
263     phi_viz = np.transpose(phi)
264     words_to_display = ~np.all(phi_viz <= viz_threshold, axis=1)
265     words_viz = [words[i] for i in range(len(words_to_display)) if
266                 words_to_display[i]]
267     phi_viz = phi_viz[words_to_display]
268     for j in range(num_topics):
269         print("Topico ", j+1)
270         valores = []
271         for i in range(len(phi_viz)):
272             valores.append({"word": words_viz[i], "phi": phi_viz[i][j]})
273         for v in sorted(valores, key=lambda x: x["phi"], reverse=True)[:10]:
274             print("    {}: {}".format(v["word"], v["phi"]))
275
276
277     fig, ax = plt.subplots()
278     heatmap = plt.pcolor(phi_viz, cmap=plt.cm.Blues, alpha=0.8)
279     plt.colorbar()
280
281     #fig.set_size_inches(8, 11)
282     ax.grid(False)
283     ax.set_frame_on(False)
284
285     ax.set_xticks(np.arange(phi_viz.shape[1]) + 0.5, minor=False)
286     ax.set_yticks(np.arange(phi_viz.shape[0]) + 0.5, minor=False)
287     ax.invert_yaxis()
288     ax.xaxis.tick_top()
289     #plt.xticks(rotation=45)
290
291     for t in ax.xaxis.get_major_ticks():
292         t.tick10n = False
293         t.tick20n = False
294     for t in ax.yaxis.get_major_ticks():
295         t.tick10n = False
296         t.tick20n = False
297
298     column_labels = words_viz #['Word ' + str(i) for i in range(1,1000)]
299     row_labels = ['Topic ' + str(i) for i in range(1,num_topics+1)]
300     ax.set_xticklabels(row_labels, minor=False)
301     ax.set_yticklabels(column_labels, minor=False)
302
303     plt.show()
304
305 def VisualizeEvolution(self, psi, timestamps):
306     xs = np.linspace(0, 1, num=1000)
307     fig, ax = plt.subplots()
308
309     print("PSI", len(psi))
310     print(psi)
311
312     for i in range(len(psi)):
313         #ys = [math.pow(1-x, psi[i][0]-1) * math.pow(x, psi[i][1]-1) / scipy.
314             special.beta(psi[i][0],psi[i][1]) for x in xs]
315         ys = [math.pow(x, psi[i][0]-1) * math.pow(1-x, psi[i][1]-1) / scipy.
316             special.beta(psi[i][0],psi[i][1]) for x in xs]
317         ax.plot(xs, ys, label='Topico ' + str(i+1))
318
319     #ax.set_xticklabels(, minor=False, )
320     plt.xticks(np.linspace(0, 1, num=len(timestamps)), [min(timestamps)]
321               + [" " for _ in range(len(timestamps) - 2)] + [max(timestamps)],
322               rotation='vertical')
323     plt.grid(True)
324
325     plt.ylabel(r'Densidade de probabilidade da função Beta, de parâmetro $
326               \Psi$')
327     plt.xlabel(r'Evolução do tempo $\rightarrow$, multialignment='
328               center')
329
330     ax.legend(loc='best', frameon=False)
331     plt.show()
332
333

```

```
327 def VisualizeHistogram(self, topics, title):
328
329     total_topics = len(topics)
330
331     lines = int(total_topics / 2)
332
333     if total_topics % 2 != 0:
334         lines += 1
335
336     fig, ax = plt.subplots(lines, 2, sharey=True)
337     fig.text(0.04, 0.5, title, va='center', rotation='vertical',
338            multialignment="center")
339
340     count = 1
341     for t in topics:
342         x_label = []
343         y_value = []
344
345         for year in sorted(t):
346             x_label.append(year)
347             y_value.append(t[year])
348
349         plt.subplot(total_topics/2 + 1, 2, count)
350
351         x = np.arange(len(x_label))
352         plt.bar(x, y_value)
353
354         if (count > total_topics - 2):
355             #plt.xticks(x, [x_label[0]] + [" " for _ in range(len(x_label) - 2)]
356                + [x_label[-1]], rotation='vertical')
357             plt.xticks(x, x_label, rotation='vertical')
358             plt.xlabel(r'Evolução do tempo $\longrightarrow$', multialignment='
359                center')
360         else:
361             plt.xticks(x, [" " * len(x_label)], )
362             #plt.title("Topico " + str(count), loc='left')
363             #plt.text(-2, 20, "Topico " + str(count))
364             plt.title("Tápico " + str(count), {'fontsize': 8}, pad=0.3)
365             plt.grid(True)
366
367         count += 1
368
369     plt.subplots_adjust(hspace=0.5)
370     plt.show()
```

Arquivo <eprint.py>.

```
1 from __future__ import print_function
2
3 import sys
4
5 def eprint(*args, **kwargs):
6     print(*args, file=sys.stderr, **kwargs)
```

Anexos

ANEXO A – Estrutura do arquivo XML de cada currículo

A listagem abaixo exhibe o arquivo DTD¹ que define todos os elementos, atributos e tipos de dados que são válidos dentro do XML de cada currículo Lattes, informando também a ordem que cada um desses itens aparece. Aparentemente os nomes foram planejados de forma clara e autoexplicativa. Para melhor visualização, os comentários desnecessários foram removidos.

A partir da análise dessa estrutura, foi selecionado um subconjunto de elementos e atributos relevantes para realização do trabalho.

```

1 <?xml version="1.0" encoding="ISO-8859-1"?>
2
3 <!ELEMENT CURRICULO-VITAE (DADOS-GERAIS, PRODUCAO-BIBLIOGRAFICA?,
4   PRODUCAO-TECNICA?, OUTRA-PRODUCAO?, DADOS-COMPLEMENTARES?)>
5 <!ATTLIST CURRICULO-VITAE
6   SISTEMA-ORIGEM-XML CDATA #REQUIRED
7   NUMERO-IDENTIFICADOR CDATA #IMPLIED
8   FORMATO-DATA-ATUALIZACAO NMTOKEN #FIXED "DDMMAAAA"
9   DATA-ATUALIZACAO CDATA #IMPLIED
10  FORMATO-HORA-ATUALIZACAO NMTOKEN #FIXED "HHMMSS"
11  HORA-ATUALIZACAO CDATA #IMPLIED
12  xmlns:lattes CDATA #IMPLIED
13 >
14 <!ELEMENT DADOS-GERAIS (RESUMO-CV?,OUTRAS-INFORMACOES-RELEVANTES? ,
15   ENDERECO?, FORMACAO-ACADEMICA-TITULACAO?, ATUACOES-PROFISSIONAIS?,
16   AREAS-DE-ATUACAO?, IDIOMAS?, PREMIOS-TITULOS?)>
17 <!ATTLIST DADOS-GERAIS
18   NOME-COMPLETO CDATA #REQUIRED
19   NOME-EM-CITACOES-BIBLIOGRAFICAS CDATA #REQUIRED
20   NACIONALIDADE CDATA #REQUIRED
21   CPF CDATA #IMPLIED
22   NUMERO-DO-PASSAPORTE CDATA #IMPLIED
23   PAIS-DE-NASCIMENTO CDATA #IMPLIED
24   UF-NASCIMENTO CDATA #IMPLIED
25   CIDADE-NASCIMENTO CDATA #IMPLIED
26   FORMATO-DATA-DE-NASCIMENTO NMTOKEN #FIXED "DDMMAAAA"
27   DATA-NASCIMENTO CDATA #IMPLIED
28   SEXO (MASCULINO | FEMININO) #REQUIRED
29   NUMERO-IDENTIDADE CDATA #IMPLIED
30   ORGAO-EMISSOR CDATA #IMPLIED
31   UF-ORGAO-EMISSOR CDATA #IMPLIED
32   FORMATO-DATA-DE-EMISSAO NMTOKEN #FIXED "DDMMAAAA"
33   DATA-DE-EMISSAO CDATA #IMPLIED
34   NOME-DO-PAI CDATA #IMPLIED
35   NOME-DA-MAE CDATA #IMPLIED
36   PERMISSAO-DE-DIVULGACAO (SIM | NAO) #REQUIRED
37   NOME-DO-ARQUIVO-DE-FOTO CDATA #IMPLIED
38   TEXTO-RESUMO-CV-RH CDATA #IMPLIED
39   OUTRAS-INFORMACOES-RELEVANTES CDATA #IMPLIED
40   DATA-FALECIMENTO CDATA #IMPLIED
41   SIGLA-PAIS-NACIONALIDADE CDATA #IMPLIED
42   PAIS-DE-NACIONALIDADE CDATA #IMPLIED
43   RACA-OU-COR CDATA #IMPLIED
44 >

```

¹ Durante a realização deste trabalho, o arquivo DTD estava disponível em <<http://lattes.cnpq.br/web/plataforma-lattes/extracao-de-dados>>, mais especificamente no link que leva a <<http://lml.cnpq.br/lmpl/Gramaticas/Curriculo/DTD/Fontes/LMPLCurriculo.DTD>>

```
42
43 <!ELEMENT RESUMO-CV EMPTY>
44 <!ATTLIST RESUMO-CV
45   TEXTO-RESUMO-CV-RH CDATA #IMPLIED
46   TEXTO-RESUMO-CV-RH-EN CDATA #IMPLIED
47 >
48 <!ELEMENT OUTRAS-INFORMACOES-RELEVANTES EMPTY>
49 <!ATTLIST OUTRAS-INFORMACOES-RELEVANTES
50   OUTRAS-INFORMACOES-RELEVANTES CDATA #IMPLIED
51 >
52
53
54 <!ELEMENT ENDERECO (ENDERECO-PROFISSIONAL?, ENDERECO-RESIDENCIAL?)>
55 <!ATTLIST ENDERECO
56   FLAG-DE-PREFERENCIA (ENDERECO_INSTITUCIONAL | ENDERECO_RESIDENCIAL) #
   IMPLIED
57 >
58 <!ELEMENT ENDERECO-PROFISSIONAL EMPTY>
59 <!ATTLIST ENDERECO-PROFISSIONAL
60   CODIGO-INSTITUICAO-EMPRESA CDATA #IMPLIED
61   NOME-INSTITUICAO-EMPRESA CDATA #IMPLIED
62   CODIGO-ORGAO CDATA #IMPLIED
63   NOME-ORGAO CDATA #IMPLIED
64   CODIGO-UNIDADE CDATA #IMPLIED
65   NOME-UNIDADE CDATA #IMPLIED
66   LOGRADOURO-COMPLEMENTO CDATA #IMPLIED
67   PAIS CDATA #IMPLIED
68   UF CDATA #IMPLIED
69   CEP CDATA #IMPLIED
70   CIDADE CDATA #IMPLIED
71   BAIRRO CDATA #IMPLIED
72   DDD CDATA #IMPLIED
73   TELEFONE CDATA #IMPLIED
74   RAMAL CDATA #IMPLIED
75   FAX CDATA #IMPLIED
76   CAIXA-POSTAL CDATA #IMPLIED
77   E-MAIL CDATA #IMPLIED
78   HOME-PAGE CDATA #IMPLIED
79 >
80 <!ELEMENT ENDERECO-RESIDENCIAL EMPTY>
81 <!ATTLIST ENDERECO-RESIDENCIAL
82   LOGRADOURO CDATA #IMPLIED
83   PAIS CDATA #IMPLIED
84   UF CDATA #IMPLIED
85   CEP CDATA #IMPLIED
86   CIDADE CDATA #IMPLIED
87   BAIRRO CDATA #IMPLIED
88   DDD CDATA #IMPLIED
89   TELEFONE CDATA #IMPLIED
90   RAMAL CDATA #IMPLIED
91   FAX CDATA #IMPLIED
92   CAIXA-POSTAL CDATA #IMPLIED
93   E-MAIL CDATA #IMPLIED
94   HOME-PAGE CDATA #IMPLIED
95 >
96 <!ELEMENT PALAVRAS-CHAVE EMPTY>
97 <!ATTLIST PALAVRAS-CHAVE
98   PALAVRA-CHAVE-1 CDATA #IMPLIED
99   PALAVRA-CHAVE-2 CDATA #IMPLIED
100  PALAVRA-CHAVE-3 CDATA #IMPLIED
101  PALAVRA-CHAVE-4 CDATA #IMPLIED
102  PALAVRA-CHAVE-5 CDATA #IMPLIED
103  PALAVRA-CHAVE-6 CDATA #IMPLIED
104 >
105 <!ELEMENT AREAS-DO-CONHECIMENTO (AREA-DO-CONHECIMENTO-1?, AREA-DO-
   CONHECIMENTO-2?, AREA-DO-CONHECIMENTO-3?)>
106 <!ELEMENT AREA-DO-CONHECIMENTO-1 EMPTY>
107 <!ATTLIST AREA-DO-CONHECIMENTO-1
108   NOME-GRANDE-AREA-DO-CONHECIMENTO (OUTROS | LINGUISTICA_LETRAS_E_ARTES |
   CIENCIAS_HUMANAS | CIENCIAS_SOCIAIS_APLICADAS | CIENCIAS_AGRARIAS |
   CIENCIAS_DA_SAUDE | ENGENHARIAS | CIENCIAS_BIOLÓGICAS |
   CIENCIAS_EXATAS_E_DA_TERRA) #IMPLIED
109   NOME-DA-AREA-DO-CONHECIMENTO CDATA #IMPLIED
```

```
110 NOME-DA-SUB-AREA-DO-CONHECIMENTO CDATA #IMPLIED
111 NOME-DA-ESPECIALIDADE CDATA #IMPLIED
112 >
113 <!ELEMENT AREA-DO-CONHECIMENTO-2 EMPTY>
114 <!ATTLIST AREA-DO-CONHECIMENTO-2
115 NOME-GRANDE-AREA-DO-CONHECIMENTO (OUTROS | LINGUISTICA_LETRAS_E_ARTES |
    CIENCIAS_HUMANAS | CIENCIAS_SOCIAIS_APLICADAS | CIENCIAS_AGRARIAS |
    CIENCIAS_DA_SAUDE | ENGENHARIAS | CIENCIAS_BIOLÓGICAS |
    CIENCIAS_EXATAS_E_DA_TERRA) #IMPLIED
116 NOME-DA-AREA-DO-CONHECIMENTO CDATA #IMPLIED
117 NOME-DA-SUB-AREA-DO-CONHECIMENTO CDATA #IMPLIED
118 NOME-DA-ESPECIALIDADE CDATA #IMPLIED
119 >
120 <!ELEMENT AREA-DO-CONHECIMENTO-3 EMPTY>
121 <!ATTLIST AREA-DO-CONHECIMENTO-3
122 NOME-GRANDE-AREA-DO-CONHECIMENTO (OUTROS | LINGUISTICA_LETRAS_E_ARTES |
    CIENCIAS_HUMANAS | CIENCIAS_SOCIAIS_APLICADAS | CIENCIAS_AGRARIAS |
    CIENCIAS_DA_SAUDE | ENGENHARIAS | CIENCIAS_BIOLÓGICAS |
    CIENCIAS_EXATAS_E_DA_TERRA) #IMPLIED
123 NOME-DA-AREA-DO-CONHECIMENTO CDATA #IMPLIED
124 NOME-DA-SUB-AREA-DO-CONHECIMENTO CDATA #IMPLIED
125 NOME-DA-ESPECIALIDADE CDATA #IMPLIED
126 >
127 <!ELEMENT SETORES-DE-ATIVIDADE EMPTY>
128 <!ATTLIST SETORES-DE-ATIVIDADE
129 SETOR-DE-ATIVIDADE-1 CDATA #IMPLIED
130 SETOR-DE-ATIVIDADE-2 CDATA #IMPLIED
131 SETOR-DE-ATIVIDADE-3 CDATA #IMPLIED
132 >
133 <!ELEMENT FORMACAO-ACADEMICA-TITULACAO (GRADUACAO*, ESPECIALIZACAO*,
    MESTRADO*, DOUTORADO*, POS-DOUTORADO*, LIVRE-DOCENCIA*, CURSO-TECNICO
    -PROFISSIONALIZANTE*, MESTRADO-PROFISSIONALIZANTE*, ENSINO-
    FUNDAMENTAL-PRIMEIRO-GRAU*, ENSINO-MEDIO-SEGUNDO-GRAU*, RESIDENCIA-
    MEDICA*, APERFEICOAMENTO*)>
134 <!ELEMENT CURSO-TECNICO-PROFISSIONALIZANTE EMPTY>
135 <!ATTLIST CURSO-TECNICO-PROFISSIONALIZANTE
136 SEQUENCIA-FORMACAO CDATA #IMPLIED
137 NIVEL CDATA #IMPLIED
138 CODIGO-INSTITUICAO CDATA #IMPLIED
139 NOME-INSTITUICAO CDATA #IMPLIED
140 CODIGO-ORGAO CDATA #IMPLIED
141 NOME-ORGAO CDATA #IMPLIED
142 CODIGO-CURSO CDATA #IMPLIED
143 NOME-CURSO CDATA #IMPLIED
144 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
145 ANO-DE-INICIO CDATA #IMPLIED
146 ANO-DE-CONCLUSAO CDATA #IMPLIED
147 FLAG-BOLSA CDATA #IMPLIED
148 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
149 NOME-AGENCIA CDATA #IMPLIED
150 NOME-CURSO-INGLES CDATA #IMPLIED
151 >
152 <!ELEMENT ENSINO-FUNDAMENTAL-PRIMEIRO-GRAU EMPTY>
153 <!ATTLIST ENSINO-FUNDAMENTAL-PRIMEIRO-GRAU
154 SEQUENCIA-FORMACAO CDATA #IMPLIED
155 NIVEL CDATA #IMPLIED
156 CODIGO-INSTITUICAO CDATA #IMPLIED
157 NOME-INSTITUICAO CDATA #IMPLIED
158 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
159 ANO-DE-INICIO CDATA #IMPLIED
160 ANO-DE-CONCLUSAO CDATA #IMPLIED
161 >
162 <!ELEMENT ENSINO-MEDIO-SEGUNDO-GRAU EMPTY>
163 <!ATTLIST ENSINO-MEDIO-SEGUNDO-GRAU
164 SEQUENCIA-FORMACAO CDATA #IMPLIED
165 NIVEL CDATA #IMPLIED
166 CODIGO-INSTITUICAO CDATA #IMPLIED
167 NOME-INSTITUICAO CDATA #IMPLIED
168 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
169 ANO-DE-INICIO CDATA #IMPLIED
170 ANO-DE-CONCLUSAO CDATA #IMPLIED
171 >
172 <!ELEMENT GRADUACAO EMPTY>
```

```
173 <!ATTLIST GRADUACAO
174 SEQUENCIA-FORMACAO CDATA #IMPLIED
175 NIVEL CDATA #IMPLIED
176 TITULO-DO-TRABALHO-DE-CONCLUSAO-DE-CURSO CDATA #IMPLIED
177 NOME-DO-ORIENTADOR CDATA #IMPLIED
178 CODIGO-INSTITUICAO CDATA #IMPLIED
179 NOME-INSTITUICAO CDATA #IMPLIED
180 CODIGO-ORGAO CDATA #IMPLIED
181 NOME-ORGAO CDATA #IMPLIED
182 CODIGO-CURSO CDATA #IMPLIED
183 NOME-CURSO CDATA #IMPLIED
184 CODIGO-AREA-CURSO CDATA #IMPLIED
185 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
186 ANO-DE-INICIO CDATA #IMPLIED
187 ANO-DE-CONCLUSAO CDATA #IMPLIED
188 FLAG-BOLSA CDATA #IMPLIED
189 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
190 NOME-AGENCIA CDATA #IMPLIED
191 NUMERO-ID-ORIENTADOR CDATA #IMPLIED
192 CODIGO-CURSO-CAPES CDATA #IMPLIED
193     TITULO-DO-TRABALHO-DE-CONCLUSAO-DE-CURSO-INGLES CDATA #IMPLIED
194     NOME-CURSO-INGLES CDATA #IMPLIED
195     FORMACAO-ACADEMICA-TITULACAO CDATA #IMPLIED
196     TIPO-GRADUACAO CDATA #IMPLIED
197     CODIGO-INSTITUICAO-GRAD CDATA #IMPLIED
198     NOME-INSTITUICAO-GRAD CDATA #IMPLIED
199     CODIGO-INSTITUICAO-OUTRA-GRAD CDATA #IMPLIED
200     NOME-INSTITUICAO-OUTRA-GRAD CDATA #IMPLIED
201     NOME-ORIENTADOR-GRAD CDATA #IMPLIED
202 >
203 <!ELEMENT APERFEICOAMENTO EMPTY>
204 <!ATTLIST APERFEICOAMENTO
205 SEQUENCIA-FORMACAO CDATA #IMPLIED
206 NIVEL CDATA #IMPLIED
207 TITULO-DA-MONOGRAFIA CDATA #IMPLIED
208 NOME-DO-ORIENTADOR CDATA #IMPLIED
209 CODIGO-INSTITUICAO CDATA #IMPLIED
210 NOME-INSTITUICAO CDATA #IMPLIED
211 CODIGO-ORGAO CDATA #IMPLIED
212 NOME-ORGAO CDATA #IMPLIED
213 CODIGO-CURSO CDATA #IMPLIED
214 NOME-CURSO CDATA #IMPLIED
215 CODIGO-AREA-CURSO CDATA #IMPLIED
216 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
217 ANO-DE-INICIO CDATA #IMPLIED
218 ANO-DE-CONCLUSAO CDATA #IMPLIED
219 FLAG-BOLSA CDATA #IMPLIED
220 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
221 NOME-AGENCIA CDATA #IMPLIED
222 CARGA-HORARIA CDATA #IMPLIED
223     TITULO-DA-MONOGRAFIA-INGLES CDATA #IMPLIED
224     NOME-CURSO-INGLES CDATA #IMPLIED
225 >
226 <!ELEMENT ESPECIALIZACAO EMPTY>
227 <!ATTLIST ESPECIALIZACAO
228 SEQUENCIA-FORMACAO CDATA #IMPLIED
229 NIVEL CDATA #IMPLIED
230 TITULO-DA-MONOGRAFIA CDATA #IMPLIED
231 NOME-DO-ORIENTADOR CDATA #IMPLIED
232 CODIGO-INSTITUICAO CDATA #IMPLIED
233 NOME-INSTITUICAO CDATA #IMPLIED
234 CODIGO-ORGAO CDATA #IMPLIED
235 NOME-ORGAO CDATA #IMPLIED
236 CODIGO-CURSO CDATA #IMPLIED
237 NOME-CURSO CDATA #IMPLIED
238 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
239 ANO-DE-INICIO CDATA #IMPLIED
240 ANO-DE-CONCLUSAO CDATA #IMPLIED
241 FLAG-BOLSA CDATA #IMPLIED
242 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
243 NOME-AGENCIA CDATA #IMPLIED
244 CARGA-HORARIA CDATA #IMPLIED
245     TITULO-DA-MONOGRAFIA-INGLES CDATA #IMPLIED
```

```
246         NOME-CURSO-INGLES CDATA #IMPLIED
247 >
248 <!ELEMENT MESTRADO (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
249     ATIVIDADE?)>
249 <!ATTLIST MESTRADO
250     SEQUENCIA-FORMACAO CDATA #IMPLIED
251     NIVEL CDATA #IMPLIED
252     CODIGO-INSTITUICAO CDATA #IMPLIED
253     NOME-INSTITUICAO CDATA #IMPLIED
254     CODIGO-ORGAO CDATA #IMPLIED
255     NOME-ORGAO CDATA #IMPLIED
256     CODIGO-CURSO CDATA #IMPLIED
257     NOME-CURSO CDATA #IMPLIED
258     CODIGO-AREA-CURSO CDATA #IMPLIED
259     STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
260     ANO-DE-INICIO CDATA #IMPLIED
261     ANO-DE-CONCLUSAO CDATA #IMPLIED
262     FLAG-BOLSA CDATA #IMPLIED
263     CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
264     NOME-AGENCIA CDATA #IMPLIED
265     ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
266     TITULO-DA-DISSERTACAO-TESE CDATA #IMPLIED
267     NOME-COMPLETO-DO-ORIENTADOR CDATA #IMPLIED
268     TIPO-MESTRADO CDATA #IMPLIED
269     NUMERO-ID-ORIENTADOR CDATA #IMPLIED
270     CODIGO-CURSO-CAPES CDATA #IMPLIED
271     TITULO-DA-DISSERTACAO-TESE-INGLES CDATA #IMPLIED
272     NOME-CURSO-INGLES CDATA #IMPLIED
273     NOME-DO-CO-ORIENTADOR CDATA #IMPLIED
274     CODIGO-INSTITUICAO-DOUT CDATA #IMPLIED
275     NOME-INSTITUICAO-DOUT CDATA #IMPLIED
276     CODIGO-INSTITUICAO-OUTRA-DOUT CDATA #IMPLIED
277     NOME-INSTITUICAO-OUTRA-DOUT CDATA #IMPLIED
278     NOME-ORIENTADOR-DOUT CDATA #IMPLIED
279 >
280 <!ELEMENT MESTRADO-PROFISSIONALIZANTE (PALAVRAS-CHAVE?, AREAS-DO-
281     CONHECIMENTO?, SETORES-DE-ATIVIDADE?)>
281 <!ATTLIST MESTRADO-PROFISSIONALIZANTE
282     SEQUENCIA-FORMACAO CDATA #IMPLIED
283     NIVEL CDATA #IMPLIED
284     CODIGO-INSTITUICAO CDATA #IMPLIED
285     NOME-INSTITUICAO CDATA #IMPLIED
286     CODIGO-ORGAO CDATA #IMPLIED
287     NOME-ORGAO CDATA #IMPLIED
288     CODIGO-CURSO CDATA #IMPLIED
289     NOME-CURSO CDATA #IMPLIED
290     CODIGO-AREA-CURSO CDATA #IMPLIED
291     STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
292     ANO-DE-INICIO CDATA #IMPLIED
293     ANO-DE-CONCLUSAO CDATA #IMPLIED
294     FLAG-BOLSA CDATA #IMPLIED
295     CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
296     NOME-AGENCIA CDATA #IMPLIED
297     ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
298     TITULO-DA-DISSERTACAO-TESE CDATA #IMPLIED
299     NOME-COMPLETO-DO-ORIENTADOR CDATA #IMPLIED
300     NUMERO-ID-ORIENTADOR CDATA #IMPLIED
301     CODIGO-CURSO-CAPES CDATA #IMPLIED
302     TITULO-DA-DISSERTACAO-TESE-INGLES CDATA #IMPLIED
303     NOME-CURSO-INGLES CDATA #IMPLIED
304     NOME-DO-CO-ORIENTADOR CDATA #IMPLIED
305 >
306 <!ELEMENT DOUTORADO (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
307     -ATIVIDADE?)>
307 <!ATTLIST DOUTORADO
308     SEQUENCIA-FORMACAO CDATA #IMPLIED
309     NIVEL CDATA #IMPLIED
310     CODIGO-INSTITUICAO CDATA #IMPLIED
311     NOME-INSTITUICAO CDATA #IMPLIED
312     CODIGO-ORGAO CDATA #IMPLIED
313     NOME-ORGAO CDATA #IMPLIED
314     CODIGO-CURSO CDATA #IMPLIED
315     NOME-CURSO CDATA #IMPLIED
```

```
316 CODIGO-AREA-CURSO CDATA #IMPLIED
317 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
318 ANO-DE-INICIO CDATA #IMPLIED
319 ANO-DE-CONCLUSAO CDATA #IMPLIED
320 FLAG-BOLSA CDATA #IMPLIED
321 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
322 NOME-AGENCIA CDATA #IMPLIED
323 ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
324 TITULO-DA-DISSERTACAO-TESE CDATA #IMPLIED
325 NOME-COMPLETO-DO-ORIENTADOR CDATA #IMPLIED
326 TIPO-DOCTORADO CDATA #IMPLIED
327 CODIGO-INSTITUICAO-DOUT CDATA #IMPLIED
328 NOME-INSTITUICAO-DOUT CDATA #IMPLIED
329 CODIGO-INSTITUICAO-OUTRA-DOUT CDATA #IMPLIED
330 NOME-INSTITUICAO-OUTRA-DOUT CDATA #IMPLIED
331 NOME-ORIENTADOR-DOUT CDATA #IMPLIED
332 NUMERO-ID-ORIENTADOR CDATA #IMPLIED
333 CODIGO-CURSO-CAPEs CDATA #IMPLIED
334 TITULO-DA-DISSERTACAO-TESE-INGLES CDATA #IMPLIED
335 NOME-CURSO-INGLES CDATA #IMPLIED
336 NOME-DO-ORIENTADOR-CO-TUTELA CDATA #IMPLIED
337 CODIGO-INSTITUICAO-OUTRA-CO-TUTELA CDATA #IMPLIED
338 CODIGO-INSTITUICAO-CO-TUTELA CDATA #IMPLIED
339 NOME-DO-ORIENTADOR-SANDUICHE CDATA #IMPLIED
340 CODIGO-INSTITUICAO-OUTRA-SANDUICHE CDATA #IMPLIED
341 CODIGO-INSTITUICAO-SANDUICHE CDATA #IMPLIED
342 NOME-DO-CO-ORIENTADOR CDATA #IMPLIED
343 >
344 <!ELEMENT RESIDENCIA-MEDICA (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?)>
345 <!ATTLIST RESIDENCIA-MEDICA
346 SEQUENCIA-FORMACAO CDATA #IMPLIED
347 NIVEL CDATA #IMPLIED
348 CODIGO-INSTITUICAO CDATA #IMPLIED
349 NOME-INSTITUICAO CDATA #IMPLIED
350 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
351 ANO-DE-INICIO CDATA #IMPLIED
352 ANO-DE-CONCLUSAO CDATA #IMPLIED
353 FLAG-BOLSA CDATA #IMPLIED
354 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
355 NOME-AGENCIA CDATA #IMPLIED
356 TITULO-DA-RESIDENCIA-MEDICA CDATA #IMPLIED
357 NUMERO-DO-REGISTRO CDATA #IMPLIED
358 TITULO-DA-RESIDENCIA-MEDICA-INGLES CDATA #IMPLIED
359 >
360 <!ELEMENT LIVRE-DOCENCIA (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?)>
361 <!ATTLIST LIVRE-DOCENCIA
362 SEQUENCIA-FORMACAO CDATA #IMPLIED
363 NIVEL CDATA #IMPLIED
364 CODIGO-INSTITUICAO CDATA #IMPLIED
365 NOME-INSTITUICAO CDATA #IMPLIED
366 ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
367 TITULO-DO-TRABALHO CDATA #IMPLIED
368 TITULO-DO-TRABALHO-INGLES CDATA #IMPLIED
369 >
370 <!ELEMENT POS-DOCTORADO (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?)>
371 <!ATTLIST POS-DOCTORADO
372 SEQUENCIA-FORMACAO CDATA #IMPLIED
373 NIVEL CDATA #IMPLIED
374 CODIGO-INSTITUICAO CDATA #IMPLIED
375 NOME-INSTITUICAO CDATA #IMPLIED
376 ANO-DE-INICIO CDATA #IMPLIED
377 ANO-DE-CONCLUSAO CDATA #IMPLIED
378 ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
379 FLAG-BOLSA CDATA #IMPLIED
380 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
381 NOME-AGENCIA CDATA #IMPLIED
382 STATUS-DO-ESTAGIO CDATA #IMPLIED
383 STATUS-DO-CURSO CDATA #IMPLIED
384 NUMERO-ID-ORIENTADOR CDATA #IMPLIED
385 CODIGO-CURSO-CAPEs CDATA #IMPLIED
```

```

386         TITULO-DO-TRABALHO CDATA #IMPLIED
387         TITULO-DO-TRABALHO-INGLES CDATA #IMPLIED
388         NOME-CURSO-INGLES CDATA #IMPLIED
389     >
390 <!-- #####
-->
391 <!-- Definicao dos elementos utilizados em Atuacao Profissional-->
392 <!-- #####
-->
393 <!ATTLIST DISCIPLINA
394     SEQUENCIA-ESPECIFICACAO CDATA #IMPLIED
395 >
396 <!ELEMENT DISCIPLINA (#PCDATA)>
397 <!ELEMENT LINHA-DE-PESQUISA (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
398     SETORES-DE-ATIVIDADE?)>
399 <!ATTLIST LINHA-DE-PESQUISA
400     SEQUENCIA-LINHA CDATA #IMPLIED
401     TITULO-DA-LINHA-DE-PESQUISA CDATA #IMPLIED
402     FLAG-LINHA-DE-PESQUISA-ATIVA (SIM | NAO) #IMPLIED
403     OBJETIVOS-LINHA-DE-PESQUISA CDATA #IMPLIED
404     TITULO-DA-LINHA-DE-PESQUISA-INGLES CDATA #IMPLIED
405     OBJETIVOS-LINHA-DE-PESQUISA-INGLES CDATA #IMPLIED
406 >
407 <!ELEMENT PROJETO-DE-PESQUISA (EQUIPE-DO-PROJETO?, FINANCIADORES-DO-
408     PROJETO?, PRODUcoes-CT-DO-PROJETO?, ORIENTACOES?)>
409 <!ATTLIST PROJETO-DE-PESQUISA
410     SEQUENCIA-PROJETO CDATA #IMPLIED
411     ANO-INICIO CDATA #IMPLIED
412     ANO-FIM CDATA #IMPLIED
413     NOME-DO-PROJETO CDATA #IMPLIED
414     SITUACAO (EM_ANDAMENTO | DESATIVADO | CONCLUIDO) #IMPLIED
415     NATUREZA (DESENVOLVIMENTO | EXTENSAO | PESQUISA | OUTRA) #IMPLIED
416     NUMERO-GRADUACAO CDATA #IMPLIED
417     NUMERO-ESPECIALIZACAO CDATA #IMPLIED
418     NUMERO-MESTRADO-ACADEMICO CDATA #IMPLIED
419     NUMERO-MESTRADO-PROF CDATA #IMPLIED
420     NUMERO-DOUTORADO CDATA #IMPLIED
421     DESCRICAO-DO-PROJETO CDATA #IMPLIED
422     IDENTIFICADOR-PROJETO CDATA #IMPLIED
423     DESCRICAO-DO-PROJETO-INGLES CDATA #IMPLIED
424     NOME-DO-PROJETO-INGLES CDATA #IMPLIED
425     FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
426     FLAG-PROJETO-CERTIFICADO (SIM | NAO) "NAO"
427     NOME-COORDENADOR-CERTIFICACAO CDATA #IMPLIED
428     FORMATO-DATA-CERTIFICACAO NMTOKEN #FIXED "DDMMAAAA"
429     DATA-CERTIFICACAO CDATA #IMPLIED
430     NUMERO_TECNICO_NIVEL_MEDIO CDATA #IMPLIED
431 >
432 <!ELEMENT EQUIPE-DO-PROJETO (INTEGRANTES-DO-PROJETO*)>
433 <!ELEMENT INTEGRANTES-DO-PROJETO (#PCDATA)>
434 <!ATTLIST INTEGRANTES-DO-PROJETO
435     NOME-COMPLETO CDATA #IMPLIED
436     NOME-PARA-CITACAO CDATA #IMPLIED
437     ORDEM-DE-INTEGRACAO CDATA #IMPLIED
438     FLAG-RESPONSAVEL (SIM | NAO) #IMPLIED
439     NRO-ID-CNPQ CDATA #IMPLIED
440 >
441 <!ELEMENT FINANCIADORES-DO-PROJETO (FINANCIADOR-DO-PROJETO+)>
442 <!ELEMENT FINANCIADOR-DO-PROJETO EMPTY>
443 <!ATTLIST FINANCIADOR-DO-PROJETO
444     SEQUENCIA-FINANCIADOR CDATA #IMPLIED
445     CODIGO-INSTITUICAO CDATA #IMPLIED
446     NOME-INSTITUICAO CDATA #IMPLIED
447     NATUREZA (BOLSA | AUXILIO_FINANCEIRO | REMUNERACAO | OUTRO | COOPERACAO
448     | NAO_INFORMADO) #IMPLIED
449 >
450 <!ELEMENT PRODUcoes-CT-DO-PROJETO (PRODUCAO-CT-DO-PROJETO+)>
451 <!ELEMENT PRODUCAO-CT-DO-PROJETO EMPTY>
452 <!ATTLIST PRODUCAO-CT-DO-PROJETO
453     SEQUENCIA-PRODUCAO-CT CDATA #IMPLIED
454     TITULO-DA-PRODUCAO-CT CDATA #IMPLIED
455     TIPO-PRODUCAO-CT CDATA #IMPLIED
456     TITULO-DA-PRODUCAO-CT-INGLES CDATA #IMPLIED

```

```

454 >
455 <!ELEMENT ORIENTACOES (ORIENTACAO+)>
456 <!ELEMENT ORIENTACAO EMPTY>
457 <!ATTLIST ORIENTACAO
458 SEQUENCIA-ORIENTACAO CDATA #IMPLIED
459 TITULO-ORIENTACAO CDATA #IMPLIED
460 TIPO-ORIENTACAO CDATA #IMPLIED
461 TITULO-ORIENTACAO-INGLES CDATA #IMPLIED
462 >
463 <!ELEMENT TREINAMENTO (#PCDATA)>
464 <!ATTLIST TREINAMENTO
465 SEQUENCIA-ESPECIFICACAO CDATA #IMPLIED
466 >
467 <!-- #####
-->
468 <!ELEMENT ATUACOES-PROFISSIONAIS (ATUACAO-PROFISSIONAL*)>
469 <!ELEMENT ATUACAO-PROFISSIONAL (VINCULOS*, ATIVIDADES-DE-DIRECAO-E-
ADMINISTRACAO?, ATIVIDADES-DE-PESQUISA-E-DESENVOLVIMENTO?, ATIVIDADES
-DE-ENSINO?, ATIVIDADES-DE-ESTAGIO?, ATIVIDADES-DE-SERVICO-TECNICO-
ESPECIALIZADO?, ATIVIDADES-DE-EXTENSAO-UNIVERSITARIA?, ATIVIDADES-DE-
TREINAMENTO-MINISTRADO?, OUTRAS-ATIVIDADES-TECNICO-CIENTIFICA?,
ATIVIDADES-DE-CONSELHO-COMISSAO-E-CONSULTORIA?, ATIVIDADES-DE-
PARTICIPACAO-EM-PROJETO?)>
470 <!ATTLIST ATUACAO-PROFISSIONAL
471 CODIGO-INSTITUICAO CDATA #IMPLIED
472 NOME-INSTITUICAO CDATA #IMPLIED
473 SEQUENCIA-ATIVIDADE CDATA #IMPLIED
474 SEQUENCIA-IMPORTANCIA CDATA #IMPLIED
475 >
476 <!--OBSERVAR QUE PARA TIPO DO VINCULO E ENQUADRAMENTO - ENTRADA LIVRE
-->
477 <!ELEMENT VINCULOS EMPTY>
478 <!ATTLIST VINCULOS
479 SEQUENCIA-HISTORICO CDATA #IMPLIED
480 TIPO-DE-VINCULO (SERVIDOR_PUBLICO_OU_CELETISTA | SERVIDOR_PUBLICO |
CELETISTA | PROFESSOR_VISITANTE | COLABORADOR |
BOLSISTA_RECEM_DOUTOR | OUTRO | LIVRE) #IMPLIED
481 ENQUADRAMENTO-FUNCIONAL (PROFESSOR_TITULAR | OUTRO | LIVRE) #IMPLIED
482 CARGA-HORARIA-SEMANAL CDATA #IMPLIED
483 FLAG-DEDICACAO-EXCLUSIVA (SIM | NAO) #IMPLIED
484 MES-INICIO CDATA #IMPLIED
485 ANO-INICIO CDATA #IMPLIED
486 MES-FIM CDATA #IMPLIED
487 ANO-FIM CDATA #IMPLIED
488 OUTRAS-INFORMACOES CDATA #IMPLIED
489 FLAG-VINCULO-EMPREGATICIO (SIM | NAO) #IMPLIED
490 OUTRO-VINCULO-INFORMADO CDATA #IMPLIED
491 OUTRO-ENQUADRAMENTO-FUNCIONAL-INFORMADO CDATA #IMPLIED
492 OUTRO-ENQUADRAMENTO-FUNCIONAL-INFORMADO-INGLES CDATA #IMPLIED
493 OUTRAS-INFORMACOES-INGLES CDATA #IMPLIED
494 >
495 <!ELEMENT ATIVIDADES-DE-DIRECAO-E-ADMINISTRACAO (DIRECAO-E-ADMINISTRACAO
+)>
496 <!ELEMENT DIRECAO-E-ADMINISTRACAO EMPTY>
497 <!ATTLIST DIRECAO-E-ADMINISTRACAO
498 SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
499 FLAG-PERODO (ANTERIOR | ATUAL) #IMPLIED
500 MES-INICIO CDATA #IMPLIED
501 ANO-INICIO CDATA #IMPLIED
502 MES-FIM CDATA #IMPLIED
503 ANO-FIM CDATA #IMPLIED
504 CODIGO-ORGAO CDATA #IMPLIED
505 NOME-ORGAO CDATA #IMPLIED
506 CODIGO-UNIDADE CDATA #IMPLIED
507 FORMATO-CARGO-OU-FUNCAO (CHEFE_DE_DEPARTAMENTO | COORDENADOR_DE_CURSO |
COORDENADOR_DE_PROGRAMA | DECANO_DE_CENTRO | DIRETOR_DE_UNIDADE |
MEMBRO_DE_COLEGIADO_SUPERIOR | MEMBRO_DE_COMISSAO_PERMANENTE |
MEMBRO_DE_COMISSAO_TEMPORARIA | MEMBRO_DE_CONSELHO_DE_CENTRO |
MEMBRO_DE_CONSELHO_DE_UNIDADE | REITOR | VICE_REITOR_OU_PRO_REITOR |
OUTRO | LIVRE) #IMPLIED
508 CARGO-OU-FUNCAO CDATA #IMPLIED
509 NOME-UNIDADE CDATA #IMPLIED
510 CARGO-OU-FUNCAO-INGLES CDATA #IMPLIED

```

```
511 >
512 <!ELEMENT ATIVIDADES-DE-PESQUISA-E-DESENVOLVIMENTO (PESQUISA-E-
    DESENVOLVIMENTO+)>
513 <!ELEMENT PESQUISA-E-DESENVOLVIMENTO (LINHA-DE-PESQUISA*)>
514 <!ATTLIST PESQUISA-E-DESENVOLVIMENTO
515     SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
516     FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
517     MES-INICIO CDATA #IMPLIED
518     ANO-INICIO CDATA #IMPLIED
519     MES-FIM CDATA #IMPLIED
520     ANO-FIM CDATA #IMPLIED
521     CODIGO-ORGAO CDATA #IMPLIED
522     NOME-ORGAO CDATA #IMPLIED
523     CODIGO-UNIDADE CDATA #IMPLIED
524     NOME-UNIDADE CDATA #IMPLIED
525 >
526 <!ELEMENT ATIVIDADES-DE-ENSINO (ENSINO*)>
527 <!ELEMENT ENSINO (DISCIPLINA*)>
528 <!ATTLIST ENSINO
529     SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
530     FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
531     TIPO-ENSINO (GRADUACAO | POS-GRADUACAO | ESPECIALIZACAO |
        APERFEICOAMENTO | ENSINO-FUNDAMENTAL | ENSINO-MEDIO | OUTRO) #
        IMPLIED
532     MES-INICIO CDATA #IMPLIED
533     ANO-INICIO CDATA #IMPLIED
534     MES-FIM CDATA #IMPLIED
535     ANO-FIM CDATA #IMPLIED
536     CODIGO-ORGAO CDATA #IMPLIED
537     NOME-ORGAO CDATA #IMPLIED
538     CODIGO-CURSO CDATA #IMPLIED
539     NOME-CURSO CDATA #IMPLIED
540     NOME-CURSO-INGLES CDATA #IMPLIED
541 >
542 <!ELEMENT ATIVIDADES-DE-ESTAGIO (ESTAGIO+)>
543 <!ELEMENT ESTAGIO EMPTY>
544 <!ATTLIST ESTAGIO
545     SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
546     FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
547     MES-INICIO CDATA #IMPLIED
548     ANO-INICIO CDATA #IMPLIED
549     MES-FIM CDATA #IMPLIED
550     ANO-FIM CDATA #IMPLIED
551     CODIGO-ORGAO CDATA #IMPLIED
552     NOME-ORGAO CDATA #IMPLIED
553     CODIGO-UNIDADE CDATA #IMPLIED
554     NOME-UNIDADE CDATA #IMPLIED
555     ESTAGIO-REALIZADO CDATA #IMPLIED
556 >
557 <!ELEMENT ATIVIDADES-DE-SERVICO-TECNICO-ESPECIALIZADO (SERVICO-TECNICO-
    ESPECIALIZADO+)>
558 <!ELEMENT SERVICO-TECNICO-ESPECIALIZADO EMPTY>
559 <!ATTLIST SERVICO-TECNICO-ESPECIALIZADO
560     SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
561     FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
562     MES-INICIO CDATA #IMPLIED
563     ANO-INICIO CDATA #IMPLIED
564     MES-FIM CDATA #IMPLIED
565     ANO-FIM CDATA #IMPLIED
566     CODIGO-ORGAO CDATA #IMPLIED
567     NOME-ORGAO CDATA #IMPLIED
568     CODIGO-UNIDADE CDATA #IMPLIED
569     NOME-UNIDADE CDATA #IMPLIED
570     SERVICO-REALIZADO CDATA #IMPLIED
571 >
572 <!ELEMENT ATIVIDADES-DE-EXTENSAO-UNIVERSITARIA (EXTENSAO-UNIVERSITARIA+)>
    >
573 <!ELEMENT EXTENSAO-UNIVERSITARIA EMPTY>
574 <!ATTLIST EXTENSAO-UNIVERSITARIA
575     SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
576     FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
577     MES-INICIO CDATA #IMPLIED
578     ANO-INICIO CDATA #IMPLIED
```

```
579 MES-FIM CDATA #IMPLIED
580 ANO-FIM CDATA #IMPLIED
581 CODIGO-ORGAO CDATA #IMPLIED
582 NOME-ORGAO CDATA #IMPLIED
583 CODIGO-UNIDADE CDATA #IMPLIED
584 NOME-UNIDADE CDATA #IMPLIED
585 ATIVIDADE-DE-EXTENSAO-REALIZADA CDATA #IMPLIED
586 >
587 <!ELEMENT ATIVIDADES-DE-TREINAMENTO-MINISTRADO (TREINAMENTO-MINISTRADO+)
    >
588 <!ELEMENT TREINAMENTO-MINISTRADO (TREINAMENTO+)>
589 <!ATTLIST TREINAMENTO-MINISTRADO
590 SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
591 FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
592 MES-INICIO CDATA #IMPLIED
593 ANO-INICIO CDATA #IMPLIED
594 MES-FIM CDATA #IMPLIED
595 ANO-FIM CDATA #IMPLIED
596 CODIGO-ORGAO CDATA #IMPLIED
597 NOME-ORGAO CDATA #IMPLIED
598 CODIGO-UNIDADE CDATA #IMPLIED
599 NOME-UNIDADE CDATA #IMPLIED
600 >
601 <!ELEMENT OUTRAS-ATIVIDADES-TECNICO-CIENTIFICA (OUTRA-ATIVIDADE-TECNICO-
    CIENTIFICA+)>
602 <!ELEMENT OUTRA-ATIVIDADE-TECNICO-CIENTIFICA EMPTY>
603 <!ATTLIST OUTRA-ATIVIDADE-TECNICO-CIENTIFICA
604 SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
605 FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
606 MES-INICIO CDATA #IMPLIED
607 ANO-INICIO CDATA #IMPLIED
608 MES-FIM CDATA #IMPLIED
609 ANO-FIM CDATA #IMPLIED
610 CODIGO-ORGAO CDATA #IMPLIED
611 NOME-ORGAO CDATA #IMPLIED
612 CODIGO-UNIDADE CDATA #IMPLIED
613 ATIVIDADE-REALIZADA CDATA #IMPLIED
614 NOME-UNIDADE CDATA #IMPLIED
615 >
616 <!ELEMENT ATIVIDADES-DE-CONSELHO-COMISSAO-E-CONSULTORIA (CONSELHO-
    COMISSAO-E-CONSULTORIA*)>
617 <!ELEMENT CONSELHO-COMISSAO-E-CONSULTORIA EMPTY>
618 <!ATTLIST CONSELHO-COMISSAO-E-CONSULTORIA
619 SEQUENCIA-FUNCAO-ATIVIDADE CDATA #IMPLIED
620 FLAG-PERIODO (ANTERIOR | ATUAL) #IMPLIED
621 MES-INICIO CDATA #IMPLIED
622 ANO-INICIO CDATA #IMPLIED
623 MES-FIM CDATA #IMPLIED
624 ANO-FIM CDATA #IMPLIED
625 CODIGO-ORGAO CDATA #IMPLIED
626 NOME-ORGAO CDATA #IMPLIED
627 CODIGO-UNIDADE CDATA #IMPLIED
628 NOME-UNIDADE CDATA #IMPLIED
629 ESPECIFICACAO CDATA #IMPLIED
630 >
631 <!ELEMENT AREAS-DE-ATUACAO (AREA-DE-ATUACAO*)>
632 <!ELEMENT AREA-DE-ATUACAO EMPTY>
633 <!ATTLIST AREA-DE-ATUACAO
634 SEQUENCIA-AREA-DE-ATUACAO CDATA #IMPLIED
635 NOME-GRANDE-AREA-DO-CONHECIMENTO (OUTROS | LINGUISTICA_LETRAS_E_ARTES |
    CIENCIAS_HUMANAS | CIENCIAS_SOCIAIS_APLICADAS | CIENCIAS_AGRARIAS |
    CIENCIAS_DA_SAUDE | ENGENHARIAS | CIENCIAS_BIOLÓGICAS |
    CIENCIAS_EXATAS_E_DA_TERRA) #IMPLIED
636 NOME-DA-AREA-DO-CONHECIMENTO CDATA #IMPLIED
637 NOME-DA-SUB-AREA-DO-CONHECIMENTO CDATA #IMPLIED
638 NOME-DA-ESPECIALIDADE CDATA #IMPLIED
639 >
640 <!ELEMENT IDIOMAS (IDIOMA*)*>
641 <!ELEMENT IDIOMA EMPTY>
642 <!ATTLIST IDIOMA
643 IDIOMA CDATA #IMPLIED
644 DESCRICAO-DO-IDIOMA CDATA #IMPLIED
645 PROFICIENCIA-DE-LEITURA (POUCO | RAZOAVELMENTE | BEM | NAO_INFORMADO)
```

```
# IMPLIED
646 PROFICIENCIA-DE-FALA (POUCO | RAZOAVELMENTE | BEM | NAO_INFORMADO) #
    IMPLIED
647 PROFICIENCIA-DE-ESCRITA (POUCO | RAZOAVELMENTE | BEM | NAO_INFORMADO)
    # IMPLIED
648 PROFICIENCIA-DE-COMPREENSAO (POUCO | RAZOAVELMENTE | BEM |
    NAO_INFORMADO) # IMPLIED
649 >
650 <!ELEMENT PREMIOS-TITULOS (PREMIO-TITULO*)>
651 <!ELEMENT PREMIO-TITULO EMPTY>
652 <!ATTLIST PREMIO-TITULO
653     NOME-DO-PREMIO-OU-TITULO CDATA # IMPLIED
654     NOME-DA-ENTIDADE-PROMOTORA CDATA # IMPLIED
655     ANO-DA-PREMIACAO CDATA # IMPLIED
656     NOME-DO-PREMIO-OU-TITULO-INGLES CDATA # IMPLIED
657 >
658 <!ELEMENT ATIVIDADES-DE-PARTICIPACAO-EM-PROJETO (PARTICIPACAO-EM-PROJETO
    *)>
659 <!ELEMENT PARTICIPACAO-EM-PROJETO (PROJETO-DE-PESQUISA*)>
660 <!ATTLIST PARTICIPACAO-EM-PROJETO
661     SEQUENCIA-FUNCAO-ATIVIDADE CDATA # IMPLIED
662     FLAG-PERiodo (ANTERIOR | ATUAL) # IMPLIED
663     MES-INICIO CDATA # IMPLIED
664     ANO-INICIO CDATA # IMPLIED
665     MES-FIM CDATA # IMPLIED
666     ANO-FIM CDATA # IMPLIED
667     CODIGO-ORGAO CDATA # IMPLIED
668     NOME-ORGAO CDATA # IMPLIED
669     CODIGO-UNIDADE CDATA # IMPLIED
670     NOME-UNIDADE CDATA # IMPLIED
671 >
672 <!-- #
    #####
    -->
673 <!--                                     -->
674 <!--             SEGMENTO DA PRODUCAO BIBLIOGRAFICA             -->
675 <!--                                     -->
676 <!--#
    #####
    -->
677 <!ELEMENT PRODUCAO-BIBLIOGRAFICA (TRABALHOS-EM-EVENTOS?, ARTIGOS-
    PUBLICADOS?, LIVROS-E-CAPITULOS?, TEXTOS-EM-JORNAIS-OU-REVISTAS?,
    DEMAIS-TIPOS-DE-PRODUCAO-BIBLIOGRAFICA?, ARTIGOS-ACEITOS-PARA-
    PUBLICACAO?)>
678 <!ELEMENT TRABALHOS-EM-EVENTOS (TRABALHO-EM-EVENTOS*)>
679 <!ATTLIST TRABALHO-EM-EVENTOS
680     SEQUENCIA-PRODUCAO CDATA # IMPLIED
681 >
682 <!ELEMENT TRABALHO-EM-EVENTOS (DADOS-BASICOS-DO-TRABALHO?, DETALHAMENTO-
    DO-TRABALHO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
683 <!ELEMENT DADOS-BASICOS-DO-TRABALHO EMPTY>
684 <!ATTLIST DADOS-BASICOS-DO-TRABALHO
685     NATUREZA (COMPLETO | RESUMO | RESUMO_EXPANDIDO) # IMPLIED
686     TITULO-DO-TRABALHO CDATA # IMPLIED
687     ANO-DO-TRABALHO CDATA # IMPLIED
688     PAIS-DO-EVENTO CDATA # IMPLIED
689     IDIOMA CDATA # IMPLIED
690     MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
691     HOME-PAGE-DO-TRABALHO CDATA # IMPLIED
692     FLAG-RELEVANCIA (SIM | NAO) "NAO"
693     DOI CDATA # IMPLIED
694     TITULO-DO-TRABALHO-INGLES CDATA # IMPLIED
695     FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
696 >
697 <!ELEMENT DETALHAMENTO-DO-TRABALHO EMPTY>
698 <!ATTLIST DETALHAMENTO-DO-TRABALHO
699     CLASSIFICACAO-DO-EVENTO (INTERNACIONAL | NACIONAL | REGIONAL | LOCAL |
    NAO_INFORMADO) "NAO_INFORMADO"
700     NOME-DO-EVENTO CDATA # IMPLIED
701     CIDADE-DO-EVENTO CDATA # IMPLIED
702     ANO-DE-REALIZACAO CDATA # IMPLIED
```

```
703 TITULO-DOS-ANAIS-OU-PROCEEDINGS CDATA #IMPLIED
704 VOLUME CDATA #IMPLIED
705 FASCICULO CDATA #IMPLIED
706 SERIE CDATA #IMPLIED
707 PAGINA-INICIAL CDATA #IMPLIED
708 PAGINA-FINAL CDATA #IMPLIED
709 ISBN CDATA #IMPLIED
710 NOME-DA-EDITORA CDATA #IMPLIED
711 CIDADE-DA-EDITORA CDATA #IMPLIED
712 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
713 >
714 <!-- ##### OBSERVACAO
##### -->
715 <!-- PARA MANTER A COMPATIBILIZACAO COM O DATACAPES (COLETA), E
HAVENDO A -->
716 <!-- INFORMACAO SOBRE O AUTOR, INCLUIR O NUMERO DO CPF PARA
BRASILEIROS -->
717 <!-- ##### OBSERVACAO
##### -->
718 <!ELEMENT AUTORES EMPTY>
719 <!ATTLIST AUTORES
720 NOME-COMPLETO-DO-AUTOR CDATA #IMPLIED
721 NOME-PARA-CITACAO CDATA #IMPLIED
722 ORDEM-DE-AUTORIA CDATA #IMPLIED
723 CPF CDATA #IMPLIED
724 NRO-ID-CNPQ CDATA #IMPLIED
725 >
726 <!ELEMENT INFORMACOES-ADICIONAIS EMPTY>
727 <!ATTLIST INFORMACOES-ADICIONAIS
728 DESCRICAO-INFORMACOES-ADICIONAIS CDATA #IMPLIED
729 DESCRICAO-INFORMACOES-ADICIONAIS-INGLES CDATA #IMPLIED
730 >
731 <!ELEMENT ARTIGOS-PUBLICADOS (ARTIGO-PUBLICADO*)>
732 <!ELEMENT ARTIGO-PUBLICADO (DADOS-BASICOS-DO-ARTIGO?, DETALHAMENTO-DO-
ARTIGO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-
DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
733 <!ATTLIST ARTIGO-PUBLICADO
734 SEQUENCIA-PRODUCAO CDATA #IMPLIED
735 ORDEM-IMPORTANCIA CDATA #IMPLIED
736 >
737 <!ELEMENT DADOS-BASICOS-DO-ARTIGO EMPTY>
738 <!ATTLIST DADOS-BASICOS-DO-ARTIGO
739 NATUREZA (COMPLETO | RESUMO | NAO_INFORMADO) #IMPLIED
740 TITULO-DO-ARTIGO CDATA #IMPLIED
741 ANO-DO-ARTIGO CDATA #IMPLIED
742 PAIS-DE-PUBLICACAO CDATA #IMPLIED
743 IDIOMA CDATA #IMPLIED
744 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
745 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
746 FLAG-RELEVANCIA (SIM | NAO) "NAO"
747 DOI CDATA #IMPLIED
748 TITULO-DO-ARTIGO-INGLES CDATA #IMPLIED
749 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
750 >
751 <!ELEMENT DETALHAMENTO-DO-ARTIGO EMPTY>
752 <!ATTLIST DETALHAMENTO-DO-ARTIGO
753 TITULO-DO-PERIODICO-OU-REVISTA CDATA #IMPLIED
754 ISSN CDATA #IMPLIED
755 VOLUME CDATA #IMPLIED
756 FASCICULO CDATA #IMPLIED
757 SERIE CDATA #IMPLIED
758 PAGINA-INICIAL CDATA #IMPLIED
759 PAGINA-FINAL CDATA #IMPLIED
760 LOCAL-DE-PUBLICACAO CDATA #IMPLIED
761 >
762 <!ELEMENT ARTIGOS-ACEITOS-PARA-PUBLICACAO (ARTIGO-ACEITO-PARA-PUBLICACAO
*)>
763 <!ELEMENT ARTIGO-ACEITO-PARA-PUBLICACAO (DADOS-BASICOS-DO-ARTIGO?,
DETALHAMENTO-DO-ARTIGO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
764 <!ATTLIST ARTIGO-ACEITO-PARA-PUBLICACAO
765 SEQUENCIA-PRODUCAO CDATA #IMPLIED
```

```
766 >
767
768 <!ELEMENT LIVROS-E-CAPITULOS (LIVROS-PUBLICADOS-OU-ORGANIZADOS?,
769     CAPITULOS-DE-LIVROS-PUBLICADOS?)>
770 <!ELEMENT LIVROS-PUBLICADOS-OU-ORGANIZADOS (LIVRO-PUBLICADO-OU-
771     ORGANIZADO*)>
772 <!ELEMENT CAPITULOS-DE-LIVROS-PUBLICADOS (CAPITULO-DE-LIVRO-PUBLICADO*)>
773 <!ELEMENT LIVRO-PUBLICADO-OU-ORGANIZADO (DADOS-BASICOS-DO-LIVRO?,
774     DETALHAMENTO-DO-LIVRO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-
775     CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
776 <ATTLIST LIVRO-PUBLICADO-OU-ORGANIZADO
777     SEQUENCIA-PRODUCAO CDATA #IMPLIED
778     TIPO (LIVRO-PUBLICADO | LIVRO-ORGANIZADO-OU-EDICAO | NAO-INFORMADO) #
779     IMPLIED
780     NATUREZA (COLETANEA | TEXTO-INTEGRAL | VERBETE | ANAIS | CATALOGO |
781     ENCICLOPEDIA | LIVRO | OUTRA | PERIODICO | NAO-INFORMADO) "
782     NAO-INFORMADO"
783     TITULO-DO-LIVRO CDATA #IMPLIED
784     ANO CDATA #IMPLIED
785     PAIS-DE-PUBLICACAO CDATA #IMPLIED
786     IDIOMA CDATA #IMPLIED
787     MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
788     FILME | HIPERTEXTO | OUTRO | VARIOS | NAO-INFORMADO) "NAO-INFORMADO"
789     HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
790     FLAG-RELEVANCIA (SIM | NAO) "NAO"
791     DOI CDATA #IMPLIED
792     TITULO-DO-LIVRO-INGLES CDATA #IMPLIED
793     FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
794 >
795 <!ELEMENT DETALHAMENTO-DO-LIVRO EMPTY>
796 <ATTLIST DETALHAMENTO-DO-LIVRO
797     NUMERO-DE-VOLUMES CDATA #IMPLIED
798     NUMERO-DE-PAGINAS CDATA #IMPLIED
799     ISBN CDATA #IMPLIED
800     NUMERO-DA-EDICAO-REVISAO CDATA #IMPLIED
801     NUMERO-DA-SERIE CDATA #IMPLIED
802     CIDADE-DA-EDITORIA CDATA #IMPLIED
803     NOME-DA-EDITORIA CDATA #IMPLIED
804 >
805 <!ELEMENT CAPITULO-DE-LIVRO-PUBLICADO (DADOS-BASICOS-DO-CAPITULO,
806     DETALHAMENTO-DO-CAPITULO, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-
807     CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
808 <!ELEMENT DADOS-BASICOS-DO-CAPITULO EMPTY>
809 <ATTLIST CAPITULO-DE-LIVRO-PUBLICADO
810     SEQUENCIA-PRODUCAO CDATA #IMPLIED
811 >
812 <ATTLIST DADOS-BASICOS-DO-CAPITULO
813     TIPO CDATA #IMPLIED
814     TITULO-DO-CAPITULO-DO-LIVRO CDATA #IMPLIED
815     ANO CDATA #IMPLIED
816     PAIS-DE-PUBLICACAO CDATA #IMPLIED
817     IDIOMA CDATA #IMPLIED
818     MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
819     FILME | HIPERTEXTO | OUTRO | VARIOS | NAO-INFORMADO) "NAO-INFORMADO"
820     HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
821     FLAG-RELEVANCIA (SIM | NAO) "NAO"
822     DOI CDATA #IMPLIED
823     TITULO-DO-CAPITULO-DO-LIVRO-INGLES CDATA #IMPLIED
824     FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
825 >
826 <!ELEMENT DETALHAMENTO-DO-CAPITULO EMPTY>
827 <ATTLIST DETALHAMENTO-DO-CAPITULO
828     TITULO-DO-LIVRO CDATA #IMPLIED
829     NUMERO-DE-VOLUMES CDATA #IMPLIED
830     PAGINA-INICIAL CDATA #IMPLIED
831     PAGINA-FINAL CDATA #IMPLIED
832     ISBN CDATA #IMPLIED
833     ORGANIZADORES CDATA #IMPLIED
834     NUMERO-DA-EDICAO-REVISAO CDATA #IMPLIED
835     NUMERO-DA-SERIE CDATA #IMPLIED
```

```
828 CIDADE-DA-EDITORA CDATA #IMPLIED
829 NOME-DA-EDITORA CDATA #IMPLIED
830 >
831 <!ELEMENT TEXTOS-EM-JORNAIS-OU-REVISTAS (TEXTO-EM-JORNAL-OU-REVISTA*)>
832 <!ELEMENT TEXTO-EM-JORNAL-OU-REVISTA (DADOS-BASICOS-DO-TEXTO?,
      DETALHAMENTO-DO-TEXTO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-
      CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
833 <!ELEMENT DADOS-BASICOS-DO-TEXTO EMPTY>
834 <!ATTLIST TEXTO-EM-JORNAL-OU-REVISTA
835 SEQUENCIA-PRODUCAO CDATA #IMPLIED
836 >
837 <!ATTLIST DADOS-BASICOS-DO-TEXTO
838 NATUREZA (JORNAL_DE_NOTICIAS | REVISTA_MAGAZINE | NAO_INFORMADO) #
      IMPLIED
839 TITULO-DO-TEXTO CDATA #IMPLIED
840 ANO-DO-TEXTO CDATA #IMPLIED
841 PAIS-DE-PUBLICACAO CDATA #IMPLIED
842 IDIOMA CDATA #IMPLIED
843 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
      FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
844 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
845 FLAG-RELEVANCIA (SIM | NAO) "NAO"
846 DOI CDATA #IMPLIED
847 TITULO-DO-TEXTO-INGLES CDATA #IMPLIED
848 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
849 >
850 <!ELEMENT DETALHAMENTO-DO-TEXTO EMPTY>
851 <!ATTLIST DETALHAMENTO-DO-TEXTO
852 TITULO-DO-JORNAL-OU-REVISTA CDATA #IMPLIED
853 ISSN CDATA #IMPLIED
854 FORMATO-DATA-DE-PUBLICACAO NMTOKEN #FIXED "DDMMAAAA"
855 DATA-DE-PUBLICACAO CDATA #IMPLIED
856 VOLUME CDATA #IMPLIED
857 PAGINA-INICIAL CDATA #IMPLIED
858 PAGINA-FINAL CDATA #IMPLIED
859 LOCAL-DE-PUBLICACAO CDATA #IMPLIED
860 >
861 <!ELEMENT DEMAIS-TIPOS-DE-PRODUCAO-BIBLIOGRAFICA (OUTRA-PRODUCAO-
      BIBLIOGRAFICA*, PARTITURA-MUSICAL*, PREFACIO-POSFACIO*, TRADUCAO*)>
862 <!ELEMENT OUTRA-PRODUCAO-BIBLIOGRAFICA (DADOS-BASICOS-DE-OUTRA-PRODUCAO
      ?, DETALHAMENTO-DE-OUTRA-PRODUCAO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-
      DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
863 <!ATTLIST OUTRA-PRODUCAO-BIBLIOGRAFICA
864 SEQUENCIA-PRODUCAO CDATA #IMPLIED
865 >
866 <!ELEMENT DADOS-BASICOS-DE-OUTRA-PRODUCAO EMPTY>
867 <!ATTLIST DADOS-BASICOS-DE-OUTRA-PRODUCAO
868 NATUREZA CDATA #IMPLIED
869 TITULO CDATA #IMPLIED
870 ANO CDATA #IMPLIED
871 PAIS-DE-PUBLICACAO CDATA #IMPLIED
872 IDIOMA CDATA #IMPLIED
873 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
      FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
874 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
875 FLAG-RELEVANCIA (SIM | NAO) "NAO"
876 DOI CDATA #IMPLIED
877 TITULO-INGLES CDATA #IMPLIED
878 NATUREZA-INGLES CDATA #IMPLIED
879 FLAG-DIVULGACAO-CIENTIFICA CDATA #IMPLIED
880 >
881 <!ELEMENT DETALHAMENTO-DE-OUTRA-PRODUCAO EMPTY>
882 <!ATTLIST DETALHAMENTO-DE-OUTRA-PRODUCAO
883 EDITORA CDATA #IMPLIED
884 CIDADE-DA-EDITORA CDATA #IMPLIED
885 NUMERO-DE-PAGINAS CDATA #IMPLIED
886 ISSN-ISBN CDATA #IMPLIED
887 >
888 <!ELEMENT PARTITURA-MUSICAL (DADOS-BASICOS-DA-PARTITURA?, DETALHAMENTO-
      DA-PARTITURA?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
      SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
889 <!ATTLIST PARTITURA-MUSICAL
890 SEQUENCIA-PRODUCAO CDATA #IMPLIED
```

```
891 >
892 <!ELEMENT DADOS-BASICOS-DA-PARTITURA EMPTY>
893 <!ATTLIST DADOS-BASICOS-DA-PARTITURA
894 NATUREZA (CANTO | CORAL | ORQUESTRA | OUTRO | NAO_INFORMADO) #IMPLIED
895 TITULO CDATA #IMPLIED
896 ANO CDATA #IMPLIED
897 PAIS-DE-PUBLICACAO CDATA #IMPLIED
898 IDIOMA CDATA #IMPLIED
899 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
900 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
901 FLAG-RELEVANCIA (SIM | NAO) "NAO"
902 DOI CDATA #IMPLIED
903 TITULO-INGLES CDATA #IMPLIED
904 >
905 <!ELEMENT DETALHAMENTO-DA-PARTITURA EMPTY>
906 <!ATTLIST DETALHAMENTO-DA-PARTITURA
907 FORMACAO-INSTRUMENTAL CDATA #IMPLIED
908 EDITORA CDATA #IMPLIED
909 CIDADE-DA-EDITORIA CDATA #IMPLIED
910 NUMERO-DE-PAGINAS CDATA #IMPLIED
911 NUMERO-DO-CATALOGO CDATA #IMPLIED
912 >
913 <!ELEMENT PREFACIO-POSFACIO (DADOS-BASICOS-DO-PREFACIO-POSFACIO?,
    DETALHAMENTO-DO-PREFACIO-POSFACIO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-
    DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
914 <!ATTLIST PREFACIO-POSFACIO
915 SEQUENCIA-PRODUCAO CDATA #IMPLIED
916 >
917 <!ELEMENT DADOS-BASICOS-DO-PREFACIO-POSFACIO EMPTY>
918 <!ATTLIST DADOS-BASICOS-DO-PREFACIO-POSFACIO
919 TIPO (PREFACIO | POSFACIO | APRESENTACAO | INTRODUCAO) #IMPLIED
920 NATUREZA (LIVRO | OUTRA | REVISTAS_OU_PERIODICOS | NAO_INFORMADO) "
    NAO_INFORMADO"
921 TITULO CDATA #IMPLIED
922 ANO CDATA #IMPLIED
923 PAIS-DE-PUBLICACAO CDATA #IMPLIED
924 IDIOMA CDATA #IMPLIED
925 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
926 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
927 FLAG-RELEVANCIA (SIM | NAO) "NAO"
928 DOI CDATA #IMPLIED
929 TITULO-INGLES CDATA #IMPLIED
930 >
931 <!ELEMENT DETALHAMENTO-DO-PREFACIO-POSFACIO EMPTY>
932 <!ATTLIST DETALHAMENTO-DO-PREFACIO-POSFACIO
933 NOME-DO-AUTOR-DA-PUBLICACAO CDATA #IMPLIED
934 TITULO-DA-PUBLICACAO CDATA #IMPLIED
935 ISSN-ISBN CDATA #IMPLIED
936 NUMERO-DA-EDICAO-REVISAO CDATA #IMPLIED
937 VOLUME CDATA #IMPLIED
938 SERIE CDATA #IMPLIED
939 FASCICULO CDATA #IMPLIED
940 EDITORA-DO-PREFACIO-POSFACIO CDATA #IMPLIED
941 CIDADE-DA-EDITORIA CDATA #IMPLIED
942 >
943 <!ELEMENT TRADUCAO (DADOS-BASICOS-DA-TRADUCAO?, DETALHAMENTO-DA-TRADUCAO
    ?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
    ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
944 <!ELEMENT DADOS-BASICOS-DA-TRADUCAO EMPTY>
945 <!ATTLIST TRADUCAO
946 SEQUENCIA-PRODUCAO CDATA #IMPLIED
947 >
948 <!ATTLIST DADOS-BASICOS-DA-TRADUCAO
949 NATUREZA (ARTIGO | LIVRO | OUTRO | NAO_INFORMADO) #IMPLIED
950 TITULO CDATA #IMPLIED
951 ANO CDATA #IMPLIED
952 PAIS-DE-PUBLICACAO CDATA #IMPLIED
953 IDIOMA CDATA #IMPLIED
954 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
955 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
```

```

956 FLAG-RELEVANCIA (SIM | NAO) "NAO"
957 DOI CDATA #IMPLIED
958 TITULO-INGLES CDATA #IMPLIED
959 >
960 <!ELEMENT DETALHAMENTO-DA-TRADUCAO EMPTY>
961 <!ATTLIST DETALHAMENTO-DA-TRADUCAO
962 NOME-DO-AUTOR-TRADUZIDO CDATA #IMPLIED
963 TITULO-DA-OBRA-ORIGINAL CDATA #IMPLIED
964 ISSN-ISBN CDATA #IMPLIED
965 IDIOMA-DA-OBRA-ORIGINAL CDATA #IMPLIED
966 EDITORA-DA-TRADUCAO CDATA #IMPLIED
967 CIDADE-DA-EDITORIA CDATA #IMPLIED
968 NUMERO-DE-PAGINAS CDATA #IMPLIED
969 NUMERO-DA-EDICAO-REVISAO CDATA #IMPLIED
970 VOLUME CDATA #IMPLIED
971 FASCICULO CDATA #IMPLIED
972 SERIE CDATA #IMPLIED
973 >
974 <!-- -->
975 <!-- -->
976 <!-- SEGMENTO DA PRODUCAO TECNICA -->
977 <!-- -->
978 <!-- -->
979 <!ELEMENT PRODUCAO-TECNICA (CULTIVAR-REGISTRADA*, SOFTWARE*, PATENTE*,
    CULTIVAR-PROTEGIDA*, DESENHO-INDUSTRIAL*, MARCA*, TOPOGRAFIA-DE-
    CIRCUITO-INTEGRADO*, PRODUTO-TECNOLOGICO*, PROCESSOS-OU-TECNICAS*,
    TRABALHO-TECNICO*, DEMAIS-TIPOS-DE-PRODUCAO-TECNICA*)>
980 <!ELEMENT REGISTRO-OU-PATENTE EMPTY>
981 <!ATTLIST REGISTRO-OU-PATENTE
982 TIPO-PATENTE (PRIVILEGIO_DE_INOVACAO_PI | MODELO_DE_UTILIDADE_MU |
983 DESENHO_INDUSTRIAL_DI | MODELO_INDUSTRIAL_MI | PATENTE_NO_EXTERIOR_PE
    |
984 PROGRAMA_DE_COMPUTADOR_PC | OUTRO_OU | OUTRO_Ou |
985 CERTIFICADO_DE_ADICAO_CA | CULTIVAR_PROTEGIDA_CTV |
    MARCA_REGISTRADA_DE_PRODUTO_MPD |
986 MARCA_REGISTRADA_DE_SERVICO_MSV | MARCA_REGISTRADA_COLETIVA_MCL |
    MARCA_REGISTRADA_DE_CERTIFICACAO_MCT |
987 TOPOGRAFIA_DE_CIRCUITO_INTEGRADO_REGISTRADA_TCI |
    MARCA_REGISTRADA_FIGURATIVA_MRF |
    MARCA_REGISTRADA_NOMINATIVA_MRN |
988 MARCA_REGISTRADA_MISTA_MRM | MARCA_REGISTRADA_TRIDIMENSIONAL_MRT
989 ) #IMPLIED
990 CODIGO-DO-REGISTRO-OU-PATENTE CDATA #IMPLIED
991 TITULO-PATENTE CDATA #IMPLIED
992 FORMATO-DATA-PEDIDO NMTOKEN #FIXED "DDMMAAAA"
993 DATA-PEDIDO-DE-DEPOSITO CDATA #IMPLIED
994 DATA-DE-PEDIDO-DE-EXAME CDATA #IMPLIED
995 DATA-DE-CONCESSAO CDATA #IMPLIED
996 INSTITUICAO-DEPOSITO-REGISTRO CDATA #IMPLIED
997 NUMERO-DEPOSITO-PCT CDATA #IMPLIED
998 FORMATO-DATA-DEPOSITO-PCT NMTOKEN #FIXED "DDMMAAAA"
999 DATA-DEPOSITO-PCT CDATA #IMPLIED
1000 NOME-DO-TITULAR CDATA #IMPLIED
1001 >
1002 <!ELEMENT SOFTWARE (DADOS-BASICOS-DO-SOFTWARE?, DETALHAMENTO-DO-SOFTWARE
    ?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
    ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1003 <!ATTLIST SOFTWARE
1004 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1005 >
1006 <!ELEMENT DADOS-BASICOS-DO-SOFTWARE EMPTY>
1007 <!ATTLIST DADOS-BASICOS-DO-SOFTWARE
1008 NATUREZA (COMPUTACIONAL | MULTIMIDIA | OUTRO | NAO_INFORMADO) #IMPLIED
1009 TITULO-DO-SOFTWARE CDATA #IMPLIED
1010 ANO CDATA #IMPLIED
1011 PAIS CDATA #IMPLIED
1012 IDIOMA CDATA #IMPLIED
1013 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1014 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1015 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1016 DOI CDATA #IMPLIED
1017 TITULO-DO-SOFTWARE-INGLES CDATA #IMPLIED

```

```
1018 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1019 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1020 >
1021 <!ELEMENT DETALHAMENTO-DO-SOFTWARE (REGISTRO-OU-PATENTE*)>
1022 <!ATTLIST DETALHAMENTO-DO-SOFTWARE
1023 FINALIDADE CDATA #IMPLIED
1024 PLATAFORMA CDATA #IMPLIED
1025 AMBIENTE CDATA #IMPLIED
1026 DISPONIBILIDADE (RESTRITA | IRRESTRITA | NAO_INFORMADO) "NAO_INFORMADO"
1027 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1028 FINALIDADE-INGLES CDATA #IMPLIED
1029 >
1030 <!ELEMENT PRODUTO-TECNOLOGICO (DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO,
    DETALHAMENTO-DO-PRODUTO-TECNOLOGICO, AUTORES*, PALAVRAS-CHAVE?, AREAS
    -DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1031 <!ATTLIST PRODUTO-TECNOLOGICO
1032 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1033 >
1034 <!ELEMENT DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO EMPTY>
1035 <!ATTLIST DADOS-BASICOS-DO-PRODUTO-TECNOLOGICO
1036 TIPO-PRODUTO (PILOTO | PROJETO | PROTOTIPO | OUTRO | NAO_INFORMADO) #
    IMPLIED
1037 NATUREZA (APARELHO | EQUIPAMENTO | FARMACOS_E_SIMILARES | INSTRUMENTO |
    OUTRA | NAO_INFORMADO) "NAO_INFORMADO"
1038 TITULO-DO-PRODUTO CDATA #IMPLIED
1039 ANO CDATA #IMPLIED
1040 PAIS CDATA #IMPLIED
1041 IDIOMA CDATA #IMPLIED
1042 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1043 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1044 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1045 DOI CDATA #IMPLIED
1046 TITULO-DO-PRODUTO-INGLES CDATA #IMPLIED
1047 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1048 >
1049 <!ELEMENT DETALHAMENTO-DO-PRODUTO-TECNOLOGICO (REGISTRO-OU-PATENTE*)>
1050 <!ATTLIST DETALHAMENTO-DO-PRODUTO-TECNOLOGICO
1051 FINALIDADE CDATA #IMPLIED
1052 DISPONIBILIDADE CDATA #IMPLIED
1053 CIDADE-DO-PRODUTO CDATA #IMPLIED
1054 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1055 FINALIDADE-INGLES CDATA #IMPLIED
1056 >
1057 <!-- NOVOS TIPOS DE REGISTRO OU PATENTE - INICIO -->
1058 <!-- O TIPO TITULARES SERA INCLUIDO NA PROXIMA VERSAO
1059 <!ELEMENT TITULARES EMPTY>
1060 <!ATTLIST TITULARES
1061 NOME CDATA #IMPLIED
1062 NOME-PARA-CITACAO CDATA #IMPLIED
1063 ORDEM CDATA #IMPLIED
1064 NRO-ID-CNPQ CDATA #IMPLIED
1065 COD_INST CDATA #IMPLIED
1066 TIPO (PESSOA | INSTITUICAO ) "PESSOA"
1067 >
1068 -->
1069
1070 <!ELEMENT PATENTE (DADOS-BASICOS-DA-PATENTE?, DETALHAMENTO-DA-PATENTE?,
    AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
    ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1071 <!ATTLIST PATENTE
1072 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1073 >
1074 <!ELEMENT DADOS-BASICOS-DA-PATENTE EMPTY>
1075 <!ATTLIST DADOS-BASICOS-DA-PATENTE
1076 TITULO CDATA #IMPLIED
1077 ANO-DESENVOLVIMENTO CDATA #IMPLIED
1078 PAIS CDATA #IMPLIED
1079 HOME-PAGE CDATA #IMPLIED
1080 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1081 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1082 TITULO-INGLES CDATA #IMPLIED
```

```
1083 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1084 >
1085 <!ELEMENT DETALHAMENTO-DA-PATENTE (REGISTRO-OU-PATENTE*)>
1086 <!ATTLIST DETALHAMENTO-DA-PATENTE
1087 FINALIDADE CDATA #IMPLIED
1088 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1089 FINALIDADE-INGLES CDATA #IMPLIED
1090 CATEGORIA CDATA #IMPLIED
1091 >
1092
1093 <!ELEMENT CULTIVAR-PROTEGIDA (DADOS-BASICOS-DA-CULTIVAR?, DETALHAMENTO-
1094 DA-CULTIVAR?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
1095 SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1096 <!ATTLIST CULTIVAR-PROTEGIDA
1097 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1098 >
1099 <!ELEMENT CULTIVAR-REGISTRADA (DADOS-BASICOS-DA-CULTIVAR?, DETALHAMENTO-
1100 DA-CULTIVAR?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
1101 SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1102 <!ATTLIST CULTIVAR-REGISTRADA
1103 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1104 >
1105 <!ELEMENT DADOS-BASICOS-DA-CULTIVAR EMPTY>
1106 <!ATTLIST DADOS-BASICOS-DA-CULTIVAR
1107 DENOMINACAO CDATA #IMPLIED
1108 ANO-SOLICITACAO CDATA #IMPLIED
1109 PAIS CDATA #IMPLIED
1110 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1111 DENOMINACAO-INGLES CDATA #IMPLIED
1112 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1113 >
1114 <!ELEMENT DETALHAMENTO-DA-CULTIVAR (REGISTRO-OU-PATENTE*)>
1115 <!ATTLIST DETALHAMENTO-DA-CULTIVAR
1116 FINALIDADE CDATA #IMPLIED
1117 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1118 FINALIDADE-INGLES CDATA #IMPLIED
1119 >
1120 <!ELEMENT DESENHO-INDUSTRIAL (DADOS-BASICOS-DO-DESENHO-INDUSTRIAL?,
1121 DETALHAMENTO-DO-DESENHO-INDUSTRIAL?, AUTORES*, PALAVRAS-CHAVE?, AREAS
1122 -DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1123 <!ATTLIST DESENHO-INDUSTRIAL
1124 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1125 >
1126 <!ELEMENT DADOS-BASICOS-DO-DESENHO-INDUSTRIAL EMPTY>
1127 <!ATTLIST DADOS-BASICOS-DO-DESENHO-INDUSTRIAL
1128 TITULO CDATA #IMPLIED
1129 ANO-DESENVOLVIMENTO CDATA #IMPLIED
1130 PAIS CDATA #IMPLIED
1131 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1132 TITULO-INGLES CDATA #IMPLIED
1133 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1134 >
1135 <!ELEMENT DETALHAMENTO-DO-DESENHO-INDUSTRIAL (REGISTRO-OU-PATENTE*)>
1136 <!ATTLIST DETALHAMENTO-DO-DESENHO-INDUSTRIAL
1137 FINALIDADE CDATA #IMPLIED
1138 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1139 FINALIDADE-INGLES CDATA #IMPLIED
1140 >
1141 <!ELEMENT MARCA (DADOS-BASICOS-DA-MARCA?, DETALHAMENTO-DA-MARCA?,
1142 AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
1143 ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1144 <!ATTLIST MARCA
1145 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1146 >
1147 <!ELEMENT DADOS-BASICOS-DA-MARCA EMPTY>
1148 <!ATTLIST DADOS-BASICOS-DA-MARCA
1149 TITULO CDATA #IMPLIED
1150 ANO-DESENVOLVIMENTO CDATA #IMPLIED
1151 PAIS CDATA #IMPLIED
1152 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1153 TITULO-INGLES CDATA #IMPLIED
```

```
1148 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1149 >
1150 <!ELEMENT DETALHAMENTO-DA-MARCA (REGISTRO-OU-PATENTE*)>
1151 <!ATTLIST DETALHAMENTO-DA-MARCA
1152 FINALIDADE CDATA #IMPLIED
1153 FINALIDADE-INGLES CDATA #IMPLIED
1154 NATUREZA CDATA #IMPLIED
1155 >
1156
1157 <!ELEMENT TOPOGRAFIA-DE-CIRCUITO-INTEGRADO (DADOS-BASICOS-DA-TOPOGRAFIA-
DE-CIRCUITO-INTEGRADO?, DETALHAMENTO-DA-TOPOGRAFIA-DE-CIRCUITO-
INTEGRADO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1158 <!ATTLIST TOPOGRAFIA-DE-CIRCUITO-INTEGRADO
1159 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1160 >
1161 <!ELEMENT DADOS-BASICOS-DA-TOPOGRAFIA-DE-CIRCUITO-INTEGRADO EMPTY>
1162 <!ATTLIST DADOS-BASICOS-DA-TOPOGRAFIA-DE-CIRCUITO-INTEGRADO
1163 TITULO CDATA #IMPLIED
1164 ANO-DESENVOLVIMENTO CDATA #IMPLIED
1165 PAIS CDATA #IMPLIED
1166 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1167 TITULO-INGLES CDATA #IMPLIED
1168 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1169 >
1170 <!ELEMENT DETALHAMENTO-DA-TOPOGRAFIA-DE-CIRCUITO-INTEGRADO (REGISTRO-OU-
PATENTE*)>
1171 <!ATTLIST DETALHAMENTO-DA-TOPOGRAFIA-DE-CIRCUITO-INTEGRADO
1172 FINALIDADE CDATA #IMPLIED
1173 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1174 FINALIDADE-INGLES CDATA #IMPLIED
1175 >
1176
1177 <!-- NOVOS TIPOS DE REGISTRO OU PATENTE - FIM -->
1178
1179 <!ELEMENT PROCESSOS-OU-TECNICAS (DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS
?, DETALHAMENTO-DO-PROCESSOS-OU-TECNICAS?, AUTORES*, PALAVRAS-CHAVE?,
AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
ADICIONAIS?)>
1180 <!ATTLIST PROCESSOS-OU-TECNICAS
1181 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1182 >
1183 <!ELEMENT DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS EMPTY>
1184 <!ATTLIST DADOS-BASICOS-DO-PROCESSOS-OU-TECNICAS
1185 NATUREZA (ANALITICA | INSTRUMENTAL | PEDAGOGICA | PROCESSUAL |
TERAPEUTICA | OUTRA | NAO_INFORMADO) #IMPLIED
1186 TITULO-DO-PROCESSO CDATA #IMPLIED
1187 ANO CDATA #IMPLIED
1188 PAIS CDATA #IMPLIED
1189 IDIOMA CDATA #IMPLIED
1190 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1191 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1192 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1193 DOI CDATA #IMPLIED
1194 TITULO-DO-PROCESSO-INGLES CDATA #IMPLIED
1195 FLAG-POTENCIAL-INOVACAO (SIM | NAO) "NAO"
1196 >
1197 <!ELEMENT DETALHAMENTO-DO-PROCESSOS-OU-TECNICAS (REGISTRO-OU-PATENTE*)>
1198 <!ATTLIST DETALHAMENTO-DO-PROCESSOS-OU-TECNICAS
1199 FINALIDADE CDATA #IMPLIED
1200 DISPONIBILIDADE CDATA #IMPLIED
1201 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1202 CIDADE-DO-PROCESSO CDATA #IMPLIED
1203 FINALIDADE-INGLES CDATA #IMPLIED
1204 >
1205 <!ELEMENT TRABALHO-TECNICO (DADOS-BASICOS-DO-TRABALHO-TECNICO?,
DETALHAMENTO-DO-TRABALHO-TECNICO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-
DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1206 <!ATTLIST TRABALHO-TECNICO
1207 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1208 >
1209 <!ELEMENT DADOS-BASICOS-DO-TRABALHO-TECNICO EMPTY>
```

```
1210 <!ATTLIST DADOS-BASICOS-DO-TRABALHO-TECNICO
1211 NATUREZA (ASSESSORIA | CONSULTORIA | PARECER | ELABORACAO_DE_PROJETO |
RELATORIO_TECNICO | SERVICOS_NA_AREA_DA_SAUDE | OUTRA |
NAO_INFORMADO) #IMPLIED
1212 TITULO-DO-TRABALHO-TECNICO CDATA #IMPLIED
1213 ANO CDATA #IMPLIED
1214 PAIS CDATA #IMPLIED
1215 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1216 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1217 IDIOMA CDATA #IMPLIED
1218 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1219 DOI CDATA #IMPLIED
1220 TITULO-DO-TRABALHO-TECNICO-INGLES CDATA #IMPLIED
1221 >
1222 <!ELEMENT DETALHAMENTO-DO-TRABALHO-TECNICO EMPTY>
1223 <!ATTLIST DETALHAMENTO-DO-TRABALHO-TECNICO
1224 FINALIDADE CDATA #IMPLIED
1225 DURACAO-EM-MESES CDATA #IMPLIED
1226 NUMERO-DE-PAGINAS CDATA #IMPLIED
1227 DISPONIBILIDADE CDATA #IMPLIED
1228 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1229 CIDADE-DO-TRABALHO CDATA #IMPLIED
1230 FINALIDADE-INGLES CDATA #IMPLIED
1231 >
1232 <!ELEMENT DEMAIS-TIPOS-DE-PRODUCAO-TECNICA (APRESENTACAO-DE-TRABALHO*,
CARTA-MAPA-OU-SIMILAR*, CURSO-DE-CURTA-DURACAO-MINISTRADO*,
DESENVOLVIMENTO-DE-MATERIAL-DIDATICO-OU-INSTRUCIONAL*, EDITORACAO*,
MANUTENCAO-DE-OBRA-ARTISTICA*, MAQUETE*, ORGANIZACAO-DE-EVENTO*,
PROGRAMA-DE-RADIO-OU-TV*, RELATORIO-DE-PESQUISA*,MIDIA-SOCIAL-WEBSITE
-BLOG*, OUTRA-PRODUCAO-TECNICA*)>
1233 <!ELEMENT APRESENTACAO-DE-TRABALHO (DADOS-BASICOS-DA-APRESENTACAO-DE-
TRABALHO?, DETALHAMENTO-DA-APRESENTACAO-DE-TRABALHO?, AUTORES*,
PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?,
INFORMACOES-ADICIONAIS?)>
1234 <!ATTLIST APRESENTACAO-DE-TRABALHO
1235 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1236 >
1237 <!ELEMENT DADOS-BASICOS-DA-APRESENTACAO-DE-TRABALHO EMPTY>
1238 <!ATTLIST DADOS-BASICOS-DA-APRESENTACAO-DE-TRABALHO
1239 NATUREZA (COMUNICACAO | CONFERENCIA | CONGRESSO | SEMINARIO | SIMPOSIO
| OUTRA | NAO_INFORMADO) #IMPLIED
1240 TITULO CDATA #IMPLIED
1241 ANO CDATA #IMPLIED
1242 PAIS CDATA #IMPLIED
1243 IDIOMA CDATA #IMPLIED
1244 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1245 DOI CDATA #IMPLIED
1246 TITULO-INGLES CDATA #IMPLIED
1247 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1248 >
1249 <!ELEMENT DETALHAMENTO-DA-APRESENTACAO-DE-TRABALHO EMPTY>
1250 <!ATTLIST DETALHAMENTO-DA-APRESENTACAO-DE-TRABALHO
1251 NOME-DO-EVENTO CDATA #IMPLIED
1252 INSTITUICAO-PROMOTORA CDATA #IMPLIED
1253 LOCAL-DA-APRESENTACAO CDATA #IMPLIED
1254 CIDADE-DA-APRESENTACAO CDATA #IMPLIED
1255 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
1256 >
1257 <!ELEMENT CARTA-MAPA-OU-SIMILAR (DADOS-BASICOS-DE-CARTA-MAPA-OU-SIMILAR
?, DETALHAMENTO-DE-CARTA-MAPA-OU-SIMILAR?, AUTORES*, PALAVRAS-CHAVE?,
AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
ADICIONAIS?)>
1258 <!ATTLIST CARTA-MAPA-OU-SIMILAR
1259 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1260 >
1261 <!ELEMENT DADOS-BASICOS-DE-CARTA-MAPA-OU-SIMILAR EMPTY>
1262 <!ATTLIST DADOS-BASICOS-DE-CARTA-MAPA-OU-SIMILAR
1263 NATUREZA (AEROFOTOGRAMA | CARTA | FOTOGRAMA | MAPA | OUTRA |
NAO_INFORMADO) #IMPLIED
1264 TITULO CDATA #IMPLIED
1265 ANO CDATA #IMPLIED
1266 PAIS CDATA #IMPLIED
```

```
1267 IDIOMA CDATA #IMPLIED
1268 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1269 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1270 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1271 DOI CDATA #IMPLIED
1272 TITULO-INGLES CDATA #IMPLIED
1273 >
1274 <!ELEMENT DETALHAMENTO-DE-CARTA-MAPA-OU-SIMILAR EMPTY>
1275 <!ATTLIST DETALHAMENTO-DE-CARTA-MAPA-OU-SIMILAR
1276 TEMA-DA-CARTA-MAPA-OU-SIMILAR CDATA #IMPLIED
1277 TECNICA-UTILIZADA CDATA #IMPLIED
1278 FINALIDADE CDATA #IMPLIED
1279 AREA-REPRESENTADA CDATA #IMPLIED
1280 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1281 FINALIDADE-INGLES CDATA #IMPLIED
1282 >
1283 <!ELEMENT CURSO-DE-CURTA-DURACAO-MINISTRADO (DADOS-BASICOS-DE-CURSOS-
    CURTA-DURACAO-MINISTRADO?, DETALHAMENTO-DE-CURSOS-CURTA-DURACAO-
    MINISTRADO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1284 <!ATTLIST CURSO-DE-CURTA-DURACAO-MINISTRADO
1285 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1286 >
1287 <!ELEMENT DADOS-BASICOS-DE-CURSOS-CURTA-DURACAO-MINISTRADO EMPTY>
1288 <!ATTLIST DADOS-BASICOS-DE-CURSOS-CURTA-DURACAO-MINISTRADO
1289 NIVEL-DO-CURSO (EXTENSAO | APERFEICOAMENTO | ESPECIALIZACAO | OUTRA |
    NAO_INFORMADO) #IMPLIED
1290 TITULO CDATA #IMPLIED
1291 ANO CDATA #IMPLIED
1292 PAIS CDATA #IMPLIED
1293 IDIOMA CDATA #IMPLIED
1294 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1295 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1296 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1297 DOI CDATA #IMPLIED
1298 TITULO-INGLES CDATA #IMPLIED
1299 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1300 >
1301 <!ELEMENT DETALHAMENTO-DE-CURSOS-CURTA-DURACAO-MINISTRADO EMPTY>
1302 <!ATTLIST DETALHAMENTO-DE-CURSOS-CURTA-DURACAO-MINISTRADO
1303 PARTICIPACAO-DOS-AUTORES (DOCENTE | ORGANIZADOR | OUTRA | NAO_INFORMADO
    ) #IMPLIED
1304 INSTITUICAO-PROMOTORA-DO-CURSO CDATA #IMPLIED
1305 LOCAL-DO-CURSO CDATA #IMPLIED
1306 CIDADE CDATA #IMPLIED
1307 DURACAO CDATA #IMPLIED
1308 UNIDADE CDATA #IMPLIED
1309 UNIDADE-INGLES CDATA #IMPLIED
1310 >
1311 <!ELEMENT DESENVOLVIMENTO-DE-MATERIAL-DIDATICO-OU-INSTRUCIONAL (DADOS-
    BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL?, DETALHAMENTO-DO-
    MATERIAL-DIDATICO-OU-INSTRUCIONAL?, AUTORES*, PALAVRAS-CHAVE?, AREAS-
    DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1312 <!ATTLIST DESENVOLVIMENTO-DE-MATERIAL-DIDATICO-OU-INSTRUCIONAL
1313 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1314 >
1315 <!ELEMENT DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL EMPTY>
1316 <!ATTLIST DADOS-BASICOS-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL
1317 NATUREZA CDATA #IMPLIED
1318 TITULO CDATA #IMPLIED
1319 ANO CDATA #IMPLIED
1320 PAIS CDATA #IMPLIED
1321 IDIOMA CDATA #IMPLIED
1322 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1323 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1324 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1325 DOI CDATA #IMPLIED
1326 TITULO-INGLES CDATA #IMPLIED
1327 NATUREZA-INGLES CDATA #IMPLIED
1328 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
```

```
1329 >
1330 <!ELEMENT DETALHAMENTO-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL EMPTY>
1331 <!ATTLIST DETALHAMENTO-DO-MATERIAL-DIDATICO-OU-INSTRUCIONAL
1332 FINALIDADE CDATA #IMPLIED
1333 FINALIDADE-INGLES CDATA #IMPLIED
1334 >
1335 <!ELEMENT EDITORACAO (DADOS-BASICOS-DE-EDITORACAO?, DETALHAMENTO-DE-
EDITORACAO?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1336 <!ATTLIST EDITORACAO
1337 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1338 >
1339 <!ELEMENT DADOS-BASICOS-DE-EDITORACAO EMPTY>
1340 <!ATTLIST DADOS-BASICOS-DE-EDITORACAO
1341 NATUREZA (LIVRO | ANAIS | CATALOGO | COLETANEA | ENCICLOPEDIA |
PERIODICO | OUTRA | NAO_INFORMADO) #IMPLIED
1342 TITULO CDATA #IMPLIED
1343 ANO CDATA #IMPLIED
1344 PAIS CDATA #IMPLIED
1345 IDIOMA CDATA #IMPLIED
1346 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1347 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1348 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1349 DOI CDATA #IMPLIED
1350 TITULO-INGLES CDATA #IMPLIED
1351 >
1352 <!ELEMENT DETALHAMENTO-DE-EDITORACAO EMPTY>
1353 <!ATTLIST DETALHAMENTO-DE-EDITORACAO
1354 NUMERO-DE-PAGINAS CDATA #IMPLIED
1355 INSTITUICAO-PROMOTORA CDATA #IMPLIED
1356 EDITORA CDATA #IMPLIED
1357 CIDADE CDATA #IMPLIED
1358 >
1359 <!ELEMENT MANUTENCAO-DE-OBRA-ARTISTICA (DADOS-BASICOS-DE-MANUTENCAO-DE-
OBRA-ARTISTICA?, DETALHAMENTO-DE-MANUTENCAO-DE-OBRA-ARTISTICA?,
AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1360 <!ATTLIST MANUTENCAO-DE-OBRA-ARTISTICA
1361 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1362 >
1363 <!ELEMENT DADOS-BASICOS-DE-MANUTENCAO-DE-OBRA-ARTISTICA EMPTY>
1364 <!ATTLIST DADOS-BASICOS-DE-MANUTENCAO-DE-OBRA-ARTISTICA
1365 TIPO (CONSERVACAO | RESTAURACAO | OUTRO | NAO_INFORMADO) #IMPLIED
1366 NATUREZA (ARQUITETURA | DESENHO | ESCULTURA | FOTOGRAFIA | GRAVURA |
OUTRA | PINTURA | NAO_INFORMADO) "NAO_INFORMADO"
1367 TITULO CDATA #IMPLIED
1368 ANO CDATA #IMPLIED
1369 PAIS CDATA #IMPLIED
1370 IDIOMA CDATA #IMPLIED
1371 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1372 DOI CDATA #IMPLIED
1373 TITULO-INGLES CDATA #IMPLIED
1374 >
1375 <!ELEMENT DETALHAMENTO-DE-MANUTENCAO-DE-OBRA-ARTISTICA EMPTY>
1376 <!ATTLIST DETALHAMENTO-DE-MANUTENCAO-DE-OBRA-ARTISTICA
1377 NOME-DA-OBRA CDATA #IMPLIED
1378 AUTOR-DA-OBRA CDATA #IMPLIED
1379 ANO-DA-OBRA CDATA #IMPLIED
1380 ACERVO (PUBLICO | PRIVADO | NAO_INFORMADO) "NAO_INFORMADO"
1381 LOCAL CDATA #IMPLIED
1382 CIDADE CDATA #IMPLIED
1383 >
1384 <!ELEMENT MAQUETE (DADOS-BASICOS-DA-MAQUETE?, DETALHAMENTO-DA-MAQUETE?,
AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1385 <!ATTLIST MAQUETE
1386 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1387 >
1388 <!ELEMENT DADOS-BASICOS-DA-MAQUETE EMPTY>
1389 <!ATTLIST DADOS-BASICOS-DA-MAQUETE
1390 TITULO CDATA #IMPLIED
1391 ANO CDATA #IMPLIED
```

```
1392 PAIS CDATA #IMPLIED
1393 IDIOMA CDATA #IMPLIED
1394 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1395 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1396 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1397 DOI CDATA #IMPLIED
1398 TITULO-INGLES CDATA #IMPLIED
1399 >
1400 <!ELEMENT DETALHAMENTO-DA-MAQUETE EMPTY>
1401 <!ATTLIST DETALHAMENTO-DA-MAQUETE
1402 FINALIDADE CDATA #IMPLIED
1403 OBJETO-REPRESENTADO CDATA #IMPLIED
1404 MATERIAL-UTILIZADO CDATA #IMPLIED
1405 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1406 FINALIDADE-INGLES CDATA #IMPLIED
1407 >
1408 <!ELEMENT ORGANIZACAO-DE-EVENTO (DADOS-BASICOS-DA-ORGANIZACAO-DE-EVENTO
    ?, DETALHAMENTO-DA-ORGANIZACAO-DE-EVENTO?, AUTORES*, PALAVRAS-CHAVE?,
    AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
    ADICIONAIS?)>
1409 <!ATTLIST ORGANIZACAO-DE-EVENTO
1410 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1411 >
1412 <!ELEMENT DADOS-BASICOS-DA-ORGANIZACAO-DE-EVENTO EMPTY>
1413 <!ATTLIST DADOS-BASICOS-DA-ORGANIZACAO-DE-EVENTO
1414 TIPO (CONCERTO | CONCURSO | CONGRESSO | EXPOSICAO | FESTIVAL | FEIRA |
    OLIMPIADA | OUTRO | NAO_INFORMADO) #IMPLIED
1415 NATUREZA (CURADORIA | MONTAGEM | MUSEOLOGIA | ORGANIZACAO |
    NAO_INFORMADO) "NAO_INFORMADO"
1416 TITULO CDATA #IMPLIED
1417 ANO CDATA #IMPLIED
1418 PAIS CDATA #IMPLIED
1419 IDIOMA CDATA #IMPLIED
1420 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1421 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1422 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1423 DOI CDATA #IMPLIED
1424 TITULO-INGLES CDATA #IMPLIED
1425 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1426 >
1427 <!ELEMENT DETALHAMENTO-DA-ORGANIZACAO-DE-EVENTO EMPTY>
1428 <!ATTLIST DETALHAMENTO-DA-ORGANIZACAO-DE-EVENTO
1429 INSTITUICAO-PROMOTORA CDATA #IMPLIED
1430 DURACAO-EM-SEMANAS CDATA #IMPLIED
1431 FLAG-EVENTO-ITINERANTE CDATA #IMPLIED
1432 FLAG-CATALOGO CDATA #IMPLIED
1433 LOCAL CDATA #IMPLIED
1434 CIDADE CDATA #IMPLIED
1435 >
1436 <!ELEMENT OUTRA-PRODUCAO-TECNICA (DADOS-BASICOS-DE-OUTRA-PRODUCAO-
    TECNICA?, DETALHAMENTO-DE-OUTRA-PRODUCAO-TECNICA?, AUTORES*, PALAVRAS
    -CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, REGISTRO-OU-
    PATENTE?, INFORMACOES-ADICIONAIS?)>
1437 <!ATTLIST OUTRA-PRODUCAO-TECNICA
1438 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1439 >
1440 <!ELEMENT DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA EMPTY>
1441 <!ATTLIST DADOS-BASICOS-DE-OUTRA-PRODUCAO-TECNICA
1442 NATUREZA CDATA #IMPLIED
1443 TITULO CDATA #IMPLIED
1444 ANO CDATA #IMPLIED
1445 PAIS CDATA #IMPLIED
1446 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1447 IDIOMA CDATA #IMPLIED
1448 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1449 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1450 DOI CDATA #IMPLIED
1451 TITULO-INGLES CDATA #IMPLIED
1452 NATUREZA-INGLES CDATA #IMPLIED
1453 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
```

```
1454 >
1455 <!ELEMENT DETALHAMENTO-DE-OUTRA-PRODUCAO-TECNICA EMPTY>
1456 <!ATTLIST DETALHAMENTO-DE-OUTRA-PRODUCAO-TECNICA
1457 FINALIDADE CDATA #IMPLIED
1458 INSTITUICAO-PROMOTORA CDATA #IMPLIED
1459 LOCAL CDATA #IMPLIED
1460 CIDADE CDATA #IMPLIED
1461 FINALIDADE-INGLES CDATA #IMPLIED
1462 >
1463 <!ELEMENT PROGRAMA-DE-RADIO-OU-TV (DADOS-BASICOS-DO-PROGRAMA-DE-RADIO-OU
- TV?, DETALHAMENTO-DO-PROGRAMA-DE-RADIO-OU-TV?, AUTORES*, PALAVRAS-
CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
ADICIONAIS?)>
1464 <!ATTLIST PROGRAMA-DE-RADIO-OU-TV
1465 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1466 >
1467 <!ELEMENT DADOS-BASICOS-DO-PROGRAMA-DE-RADIO-OU-TV EMPTY>
1468 <!ATTLIST DADOS-BASICOS-DO-PROGRAMA-DE-RADIO-OU-TV
1469 NATUREZA (ENTREVISTA | MESA_REDONDA | COMENTARIO | PROGRAMA | OUTRA |
NAO_INFORMADO) #IMPLIED
1470 TITULO CDATA #IMPLIED
1471 ANO CDATA #IMPLIED
1472 PAIS CDATA #IMPLIED
1473 IDIOMA CDATA #IMPLIED
1474 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1475 DOI CDATA #IMPLIED
1476 TITULO-INGLES CDATA #IMPLIED
1477 HOME-PAGE CDATA #IMPLIED
1478 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1479 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1480 >
1481 <!ELEMENT DETALHAMENTO-DO-PROGRAMA-DE-RADIO-OU-TV EMPTY>
1482 <!ATTLIST DETALHAMENTO-DO-PROGRAMA-DE-RADIO-OU-TV
1483 EMISSORA CDATA #IMPLIED
1484 TEMA CDATA #IMPLIED
1485 FORMATO-DATA-DA-APRESENTACAO NMTOKEN #FIXED "DDMMAAAA"
1486 DATA-DA-APRESENTACAO CDATA #IMPLIED
1487 DURACAO-EM-MINUTOS CDATA #IMPLIED
1488 CIDADE CDATA #IMPLIED
1489 VEICULO-DE-DIVULGACAO CDATA #IMPLIED
1490 >
1491 <!ELEMENT RELATORIO-DE-PESQUISA (DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA
?, DETALHAMENTO-DO-RELATORIO-DE-PESQUISA?, AUTORES*, PALAVRAS-CHAVE?,
AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
ADICIONAIS?)>
1492 <!ATTLIST RELATORIO-DE-PESQUISA
1493 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1494 >
1495 <!ELEMENT DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA EMPTY>
1496 <!ATTLIST DADOS-BASICOS-DO-RELATORIO-DE-PESQUISA
1497 TITULO CDATA #IMPLIED
1498 ANO CDATA #IMPLIED
1499 PAIS CDATA #IMPLIED
1500 IDIOMA CDATA #IMPLIED
1501 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1502 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
1503 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1504 DOI CDATA #IMPLIED
1505 TITULO-INGLES CDATA #IMPLIED
1506 >
1507 <!ELEMENT DETALHAMENTO-DO-RELATORIO-DE-PESQUISA EMPTY>
1508 <!ATTLIST DETALHAMENTO-DO-RELATORIO-DE-PESQUISA
1509 NOME-DO-PROJETO CDATA #IMPLIED
1510 NUMERO-DE-PAGINAS CDATA #IMPLIED
1511 DISPONIBILIDADE CDATA #IMPLIED
1512 INSTITUICAO-FINANCIADORA CDATA #IMPLIED
1513 >
1514
1515 <!ELEMENT MIDIA-SOCIAL-WEBSITE-BLOG (DADOS-BASICOS-DA-MIDIA-SOCIAL-
WEBSITE-BLOG?,
1516 DETALHAMENTO-DA-MIDIA-SOCIAL-WEBSITE-BLOG?,
```

```
1517 AUTORES*,
1518 PALAVRAS-CHAVE?,
1519 AREAS-DO-CONHECIMENTO?,
1520 SETORES-DE-ATIVIDADE?,
1521 INFORMACOES-ADICIONAIS?)>
1522 <! ATTLIST MIDIA-SOCIAL-WEBSITE-BLOG
1523 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1524 >
1525 <! ELEMENT DADOS-BASICOS-DA-MIDIA-SOCIAL-WEBSITE-BLOG EMPTY>
1526 <! ATTLIST DADOS-BASICOS-DA-MIDIA-SOCIAL-WEBSITE-BLOG
1527 NATUREZA (REDE_SOCIAL | FORUM | BLOG | SITE) #IMPLIED
1528 NATUREZA-INGLES CDATA #IMPLIED
1529 TITULO CDATA #IMPLIED
1530 TITULO-INGLES CDATA #IMPLIED
1531 ANO CDATA #IMPLIED
1532 PAIS CDATA #IMPLIED
1533 IDIOMA CDATA #IMPLIED
1534 HOME-PAGE CDATA #IMPLIED
1535 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1536 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1537 >
1538 <! ELEMENT DETALHAMENTO-DA-MIDIA-SOCIAL-WEBSITE-BLOG EMPTY>
1539 <! ATTLIST DETALHAMENTO-DA-MIDIA-SOCIAL-WEBSITE-BLOG
1540 TEMA CDATA #IMPLIED
1541 >
1542
1543 <!-- -->
1544 <!-- -->
1545 <!-- SEGMENTO DE OUTRA PRODUCAO -->
1546 <!-- -->
1547 <!-- -->
1548 <! ELEMENT OUTRA-PRODUCAO (PRODUCAO-ARTISTICA-CULTURAL*, ORIENTACOES-
1549 CONCLUIDAS*, DEMAIS-TRABALHOS*)>
1550 <! ELEMENT PRODUCAO-ARTISTICA-CULTURAL (APRESENTACAO-DE-OBRA-ARTISTICA*,
1551 APRESENTACAO-EM-RADIO-OU-TV*, ARRANJO-MUSICAL*, COMPOSICAO-MUSICAL*,
1552 CURSO-DE-CURTA-DURACAO*, OBRA-DE-ARTES-VISUAIS*, OUTRA-PRODUCAO-
1553 ARTISTICA-CULTURAL*, SONOPLASTIA*, ARTES-CENICAS*, ARTES-VISUAIS*,
1554 MUSICA*)>
1555 <! ELEMENT APRESENTACAO-DE-OBRA-ARTISTICA (DADOS-BASICOS-DA-APRESENTACAO-
1556 DE-OBRA-ARTISTICA?, DETALHAMENTO-DA-APRESENTACAO-DE-OBRA-ARTISTICA?,
1557 AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
1558 ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1559 <! ATTLIST APRESENTACAO-DE-OBRA-ARTISTICA
1560 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1561 >
1562 <! ELEMENT DADOS-BASICOS-DA-APRESENTACAO-DE-OBRA-ARTISTICA EMPTY>
1563 <! ATTLIST DADOS-BASICOS-DA-APRESENTACAO-DE-OBRA-ARTISTICA
1564 NATUREZA (COREOGRAFICA | LITERARIA | MUSICAL | TEATRAL | OUTRA |
1565 NAO_INFORMADO) #IMPLIED
1566 TITULO CDATA #IMPLIED
1567 ANO CDATA #IMPLIED
1568 PAIS CDATA #IMPLIED
1569 IDIOMA CDATA #IMPLIED
1570 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
1571 FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1572 HOME-PAGE CDATA #IMPLIED
1573 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1574 DOI CDATA #IMPLIED
1575 TITULO-INGLES CDATA #IMPLIED
1576 >
1577 <! ELEMENT DETALHAMENTO-DA-APRESENTACAO-DE-OBRA-ARTISTICA EMPTY>
1578 <! ATTLIST DETALHAMENTO-DA-APRESENTACAO-DE-OBRA-ARTISTICA
1579 TIPO-DE-EVENTO (CONCERTO | CONCURSO | FESTIVAL | GRAVACAO | RECITAL |
1580 OUTRO | NAO_INFORMADO) #IMPLIED
1581 ATIVIDADE-DOS-AUTORES (CANTO | CRIACAO | DANCA | DIRECAO | ENCENACAO |
1582 INSTRUMENTO_MUSICAL | REGENCIA | ROTEIRO | OUTRA | VARIAS |
1583 NAO_INFORMADO) #IMPLIED
1584 FLAG-INEDITISMO-DA-OBRA CDATA #IMPLIED
1585 PREMIACAO CDATA #IMPLIED
1586 OBRA-DE-REFERENCIA CDATA #IMPLIED
1587 AUTOR-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1588 ANO-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1589 DURACAO-EM-MINUTOS CDATA #IMPLIED
```

```
1577 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1578 LOCAL-DO-EVENTO CDATA #IMPLIED
1579 CIDADE CDATA #IMPLIED
1580 >
1581 <!ELEMENT APRESENTACAO-EM-RADIO-OU-TV (DADOS-BASICOS-DA-APRESENTACAO-EM-
      RADIO-OU-TV?, DETALHAMENTO-DA-APRESENTACAO-EM-RADIO-OU-TV?, AUTORES*,
      PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?,
      INFORMACOES-ADICIONAIS?)>
1582 <!ATTLIST APRESENTACAO-EM-RADIO-OU-TV
1583 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1584 >
1585 <!ELEMENT DADOS-BASICOS-DA-APRESENTACAO-EM-RADIO-OU-TV EMPTY>
1586 <!ATTLIST DADOS-BASICOS-DA-APRESENTACAO-EM-RADIO-OU-TV
1587 NATUREZA (DANCA | MUSICA | TEATRO | OUTRA | NAO_INFORMADO) #IMPLIED
1588 TITULO CDATA #IMPLIED
1589 ANO CDATA #IMPLIED
1590 PAIS CDATA #IMPLIED
1591 IDIOMA CDATA #IMPLIED
1592 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1593 DOI CDATA #IMPLIED
1594 TITULO-INGLES CDATA #IMPLIED
1595 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1596 >
1597 <!ELEMENT DETALHAMENTO-DA-APRESENTACAO-EM-RADIO-OU-TV EMPTY>
1598 <!ATTLIST DETALHAMENTO-DA-APRESENTACAO-EM-RADIO-OU-TV
1599 EMISSORA CDATA #IMPLIED
1600 FORMATO-DATA-DE-APRESENTACAO NMTOKEN #FIXED "DDMMAAAA"
1601 DATA-DE-APRESENTACAO CDATA #IMPLIED
1602 DURACAO-EM-MINUTOS CDATA #IMPLIED
1603 CIDADE CDATA #IMPLIED
1604 >
1605 <!ELEMENT ARRANJO-MUSICAL (DADOS-BASICOS-DO-ARRANJO-MUSICAL?,
      DETALHAMENTO-DO-ARRANJO-MUSICAL?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO
      -CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1606 <!ATTLIST ARRANJO-MUSICAL
1607 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1608 >
1609 <!ELEMENT DADOS-BASICOS-DO-ARRANJO-MUSICAL EMPTY>
1610 <!ATTLIST DADOS-BASICOS-DO-ARRANJO-MUSICAL
1611 NATUREZA (CANTO | CORAL | ORQUESTRA | OUTRA | NAO_INFORMADO) #IMPLIED
1612 TITULO CDATA #IMPLIED
1613 ANO CDATA #IMPLIED
1614 PAIS CDATA #IMPLIED
1615 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
      FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1616 HOME-PAGE CDATA #IMPLIED
1617 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1618 DOI CDATA #IMPLIED
1619 TITULO-INGLES CDATA #IMPLIED
1620 >
1621 <!ELEMENT DETALHAMENTO-DO-ARRANJO-MUSICAL EMPTY>
1622 <!ATTLIST DETALHAMENTO-DO-ARRANJO-MUSICAL
1623 AUTOR-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1624 ANO-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1625 FORMACAO-INSTRUMENTAL CDATA #IMPLIED
1626 REGISTRO-DE-DIREITO-AUTORAL CDATA #IMPLIED
1627 PREMIACAO CDATA #IMPLIED
1628 >
1629 <!ELEMENT COMPOSICAO-MUSICAL (DADOS-BASICOS-DA-COMPOSICAO-MUSICAL?,
      DETALHAMENTO-DA-COMPOSICAO-MUSICAL?, AUTORES*, PALAVRAS-CHAVE?, AREAS
      -DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1630 <!ATTLIST COMPOSICAO-MUSICAL
1631 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1632 >
1633 <!ELEMENT DADOS-BASICOS-DA-COMPOSICAO-MUSICAL EMPTY>
1634 <!ATTLIST DADOS-BASICOS-DA-COMPOSICAO-MUSICAL
1635 NATUREZA (CANTO | CORAL | ORQUESTRA | OUTRA | NAO_INFORMADO) #IMPLIED
1636 TITULO CDATA #IMPLIED
1637 ANO CDATA #IMPLIED
1638 PAIS CDATA #IMPLIED
1639 IDIOMA CDATA #IMPLIED
1640 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
      FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
```

```
1641 HOME-PAGE CDATA #IMPLIED
1642 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1643 DOI CDATA #IMPLIED
1644 TITULO-INGLES CDATA #IMPLIED
1645 >
1646 <!ELEMENT DETALHAMENTO-DA-COMPOSICAO-MUSICAL EMPTY>
1647 <!ATTLIST DETALHAMENTO-DA-COMPOSICAO-MUSICAL
1648 FORMACAO-INSTRUMENTAL CDATA #IMPLIED
1649 NUMERO-DE-PAGINAS CDATA #IMPLIED
1650 REGISTRO-DE-DIREITO-AUTORAL CDATA #IMPLIED
1651 PREMIACAO CDATA #IMPLIED
1652 >
1653 <!ELEMENT CURSO-DE-CURTA-DURACAO (DADOS-BASICOS-DO-CURSO-DE-CURTA-
    DURACAO?, DETALHAMENTO-DO-CURSO-DE-CURTA-DURACAO?, AUTORES*, PALAVRAS
    -CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
    ADICIONAIS?)>
1654 <!ATTLIST CURSO-DE-CURTA-DURACAO
1655 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1656 >
1657 <!ELEMENT DADOS-BASICOS-DO-CURSO-DE-CURTA-DURACAO EMPTY>
1658 <!ATTLIST DADOS-BASICOS-DO-CURSO-DE-CURTA-DURACAO
1659 NATUREZA (EXTENSAO | APERFEICOAMENTO | ESPECIALIZACAO | OUTRA |
    NAO_INFORMADO) #IMPLIED
1660 TITULO CDATA #IMPLIED
1661 ANO CDATA #IMPLIED
1662 PAIS CDATA #IMPLIED
1663 IDIOMA CDATA #IMPLIED
1664 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1665 HOME-PAGE CDATA #IMPLIED
1666 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1667 DOI CDATA #IMPLIED
1668 TITULO-INGLES CDATA #IMPLIED
1669 >
1670 <!ELEMENT DETALHAMENTO-DO-CURSO-DE-CURTA-DURACAO EMPTY>
1671 <!ATTLIST DETALHAMENTO-DO-CURSO-DE-CURTA-DURACAO
1672 DURACAO CDATA #IMPLIED
1673 UNIDADE (SEMANAS | DIAS | HORAS | NAO_INFORMADO) "NAO_INFORMADO"
1674 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1675 LOCAL-DO-EVENTO CDATA #IMPLIED
1676 CIDADE CDATA #IMPLIED
1677 >
1678 <!ELEMENT OBRA-DE-ARTES-VISUAIS (DADOS-BASICOS-DA-OBRA-DE-ARTES-VISUAIS
    ?, DETALHAMENTO-DA-OBRA-DE-ARTES-VISUAIS?, AUTORES*, PALAVRAS-CHAVE?,
    AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-
    ADICIONAIS?)>
1679 <!ELEMENT DADOS-BASICOS-DA-OBRA-DE-ARTES-VISUAIS EMPTY>
1680 <!ATTLIST OBRA-DE-ARTES-VISUAIS
1681 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1682 >
1683 <!ATTLIST DADOS-BASICOS-DA-OBRA-DE-ARTES-VISUAIS
1684 NATUREZA (CINEMA | DESENHO | ESCULTURA | FOTOGRAFIA | GRAVURA |
    INSTALACAO | PINTURA | TELEVISAO | VIDEO | OUTRA) #IMPLIED
1685 TITULO CDATA #IMPLIED
1686 ANO CDATA #IMPLIED
1687 PAIS CDATA #IMPLIED
1688 IDIOMA CDATA #IMPLIED
1689 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1690 HOME-PAGE CDATA #IMPLIED
1691 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1692 DOI CDATA #IMPLIED
1693 TITULO-INGLES CDATA #IMPLIED
1694 >
1695 <!ELEMENT DETALHAMENTO-DA-OBRA-DE-ARTES-VISUAIS EMPTY>
1696 <!ATTLIST DETALHAMENTO-DA-OBRA-DE-ARTES-VISUAIS
1697 MATERIAL-EMPREGADO CDATA #IMPLIED
1698 TIPO-DE-EVENTO (APRESENTACAO | CONCURSO | CRIACAO | EXPOSICAO |
    FESTIVAL | OUTRO | NAO_INFORMADO) #IMPLIED
1699 EVENTO CDATA #IMPLIED
1700 PREMIACAO CDATA #IMPLIED
1701 ACERVO (PUBLICO | PRIVADO | NAO_INFORMADO) "NAO_INFORMADO"
1702 INSTITUICAO-PROMOTORA CDATA #IMPLIED
```

```
1703 >
1704 <!ELEMENT OUTRA-PRODUCAO-ARTISTICA-CULTURAL (DADOS-BASICOS-DE-OUTRA-
    PRODUCAO-ARTISTICA-CULTURAL?, DETALHAMENTO-DE-OUTRA-PRODUCAO-
    ARTISTICA-CULTURAL?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO
    ?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1705 <!ATTLIST OUTRA-PRODUCAO-ARTISTICA-CULTURAL
1706 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1707 >
1708 <!ELEMENT DADOS-BASICOS-DE-OUTRA-PRODUCAO-ARTISTICA-CULTURAL EMPTY>
1709 <!ATTLIST DADOS-BASICOS-DE-OUTRA-PRODUCAO-ARTISTICA-CULTURAL
1710 NATUREZA CDATA #IMPLIED
1711 TITULO CDATA #IMPLIED
1712 ANO CDATA #IMPLIED
1713 PAIS CDATA #IMPLIED
1714 IDIOMA CDATA #IMPLIED
1715 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1716 HOME-PAGE CDATA #IMPLIED
1717 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1718 DOI CDATA #IMPLIED
1719 TITULO-INGLES CDATA #IMPLIED
1720 NATUREZA-INGLES CDATA #IMPLIED
1721 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1722 >
1723 <!ELEMENT DETALHAMENTO-DE-OUTRA-PRODUCAO-ARTISTICA-CULTURAL EMPTY>
1724 <!ATTLIST DETALHAMENTO-DE-OUTRA-PRODUCAO-ARTISTICA-CULTURAL
1725 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1726 LOCAL-DO-EVENTO CDATA #IMPLIED
1727 CIDADE CDATA #IMPLIED
1728 EXPOSICAO CDATA #IMPLIED
1729 PREMIACAO CDATA #IMPLIED
1730 >
1731 <!ELEMENT SONOPLASTIA (DADOS-BASICOS-DE-SONOPLASTIA?, DETALHAMENTO-DE-
    SONOPLASTIA?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1732 <!ATTLIST SONOPLASTIA
1733 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1734 >
1735 <!ELEMENT DADOS-BASICOS-DE-SONOPLASTIA EMPTY>
1736 <!ATTLIST DADOS-BASICOS-DE-SONOPLASTIA
1737 NATUREZA (CINEMA | MUSICA | RADIO | TELEVISAO | TEATRO | OUTRA) #
    IMPLIED
1738 TITULO CDATA #IMPLIED
1739 ANO CDATA #IMPLIED
1740 PAIS CDATA #IMPLIED
1741 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1742 HOME-PAGE CDATA #IMPLIED
1743 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1744 DOI CDATA #IMPLIED
1745 TITULO-INGLES CDATA #IMPLIED
1746 >
1747 <!ELEMENT DETALHAMENTO-DE-SONOPLASTIA EMPTY>
1748 <!ATTLIST DETALHAMENTO-DE-SONOPLASTIA
1749 FINALIDADE CDATA #IMPLIED
1750 PREMIACAO CDATA #IMPLIED
1751 >
1752 <!-- INICIO - novos tipos de producao artistica - novembro de 2011 -->
1753 <!ELEMENT ARTES-CENICAS (DADOS-BASICOS-DE-ARTES-CENICAS?, DETALHAMENTO-
    DE-ARTES-CENICAS?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1754 <!ATTLIST ARTES-CENICAS
1755 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1756 >
1757 <!ELEMENT DADOS-BASICOS-DE-ARTES-CENICAS EMPTY>
1758 <!ATTLIST DADOS-BASICOS-DE-ARTES-CENICAS
1759 NATUREZA (AUDIOVISUAL | CIRCENSE | COREOGRAFICA | DIVERSAS | OPERISTICA
    | PERFORMATICA | RADIALISTICA | TEATRAL | OUTRA ) #IMPLIED
1760 TITULO CDATA #IMPLIED
1761 ANO CDATA #IMPLIED
1762 PAIS CDATA #IMPLIED
1763 IDIOMA CDATA #IMPLIED
1764 FLAG-RELEVANCIA (SIM | NAO) "NAO"
```

```
1765 TITULO-INGLES CDATA #IMPLIED
1766 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1767 HOME-PAGE CDATA #IMPLIED
1768 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1769 >
1770 <!ELEMENT DETALHAMENTO-DE-ARTES-CENICAS EMPTY>
1771 <!ATTLIST DETALHAMENTO-DE-ARTES-CENICAS
1772 TIPO-DE-EVENTO CDATA #IMPLIED
1773 ATIVIDADE-DOS-AUTORES CDATA #IMPLIED
1774 DATA-ESTREIA CDATA #IMPLIED
1775 DATA-ENCERRAMENTO CDATA #IMPLIED
1776 LOCAL-DE-ESTREIA CDATA #IMPLIED
1777 PREMIACAO CDATA #IMPLIED
1778 INSTITUICAO-PROMOTORA-DO-PREMIO CDATA #IMPLIED
1779 OBRA-DE-REFERENCIA CDATA #IMPLIED
1780 AUTOR-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1781 ANO-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1782 DURACAO CDATA #IMPLIED
1783 TEMPORADA CDATA #IMPLIED
1784 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1785 LOCAL-DO-EVENTO CDATA #IMPLIED
1786 CIDADE-DO-EVENTO CDATA #IMPLIED
1787 FLAG-ITINERANTE CDATA #IMPLIED
1788 >
1789
1790 <!ELEMENT ARTES-VISUAIS (DADOS-BASICOS-DE-ARTES-VISUAIS?, DETALHAMENTO-
    DE-ARTES-VISUAIS?, AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1791 <!ATTLIST ARTES-VISUAIS
1792 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1793 >
1794 <!ELEMENT DADOS-BASICOS-DE-ARTES-VISUAIS EMPTY>
1795 <!ATTLIST DADOS-BASICOS-DE-ARTES-VISUAIS
1796 NATUREZA (INTERVENCAO_URBANA | LIVRO_DE_ARTISTA | PERFORMANCE | PINTURA
    | PROGRAMACAO_VISUAL | VIDEO | WEBART | ANIMACAO | INSTALACAO |
1797 COMPUTACAO_GRAFICA | DESENHO | DIVERSAS | ESCULTURA | FILME |
    FOTOGRAFIA | GRAVURA | ILUSTRACAO | OUTRA ) #IMPLIED
1798 TITULO CDATA #IMPLIED
1799 ANO CDATA #IMPLIED
1800 PAIS CDATA #IMPLIED
1801 IDIOMA CDATA #IMPLIED
1802 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1803 TITULO-INGLES CDATA #IMPLIED
1804 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1805 HOME-PAGE CDATA #IMPLIED
1806 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1807 >
1808 <!ELEMENT DETALHAMENTO-DE-ARTES-VISUAIS EMPTY>
1809 <!ATTLIST DETALHAMENTO-DE-ARTES-VISUAIS
1810 PREMIACAO CDATA #IMPLIED
1811 ATIVIDADE-DOS-AUTORES CDATA #IMPLIED
1812 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1813 LOCAL-DO-EVENTO CDATA #IMPLIED
1814 CIDADE-DO-EVENTO CDATA #IMPLIED
1815 TEMPORADA CDATA #IMPLIED
1816 >
1817
1818 <!ELEMENT MUSICA (DADOS-BASICOS-DA-MUSICA?, DETALHAMENTO-DA-MUSICA?,
    AUTORES*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
    ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1819 <!ATTLIST MUSICA
1820 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1821 >
1822 <!ELEMENT DADOS-BASICOS-DA-MUSICA EMPTY>
1823 <!ATTLIST DADOS-BASICOS-DA-MUSICA
1824 NATUREZA (APRESENTACAO_DE_OBRA | ARRANJO | AUDIOVISUAL | COMPOSICAO |
    DIVERSAS | INTERPRETACAO | PUBLICACAO_DE_PARTITURA |
    REGISTRO_FONOGRAFICO | TRILHA_SONORA | OUTRA ) #IMPLIED
1825 TITULO CDATA #IMPLIED
1826 ANO CDATA #IMPLIED
1827 PAIS CDATA #IMPLIED
```

```
1828 IDIOMA CDATA #IMPLIED
1829 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
    FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
1830 HOME-PAGE CDATA #IMPLIED
1831 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1832 TITULO-INGLES CDATA #IMPLIED
1833 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
1834 >
1835 <!ELEMENT DETALHAMENTO-DA-MUSICA EMPTY>
1836 <!ATTLIST DETALHAMENTO-DA-MUSICA
1837 TIPO-DE-EVENTO CDATA #IMPLIED
1838 ATIVIDADE-DOS-AUTORES CDATA #IMPLIED
1839 FORMACAO-INSTRUMENTAL CDATA #IMPLIED
1840 FLAG-INEDITISMO-DA-OBRA CDATA #IMPLIED
1841 DATA-ESTREIA CDATA #IMPLIED
1842 DATA-ENCERRAMENTO CDATA #IMPLIED
1843 LOCAL-DE-ESTREIA CDATA #IMPLIED
1844 PREMIACAO CDATA #IMPLIED
1845 INSTITUICAO-PROMOTORA-DO-PREMIO CDATA #IMPLIED
1846 OBRA-DE-REFERENCIA CDATA #IMPLIED
1847 AUTOR-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1848 ANO-DA-OBRA-DE-REFERENCIA CDATA #IMPLIED
1849 DURACAO CDATA #IMPLIED
1850 TEMPORADA CDATA #IMPLIED
1851 INSTITUICAO-PROMOTORA-DO-EVENTO CDATA #IMPLIED
1852 LOCAL-DO-EVENTO CDATA #IMPLIED
1853 CIDADE-DO-EVENTO CDATA #IMPLIED
1854 >
1855
1856 <!-- FIM - novos tipos de producao artistica - novembro de 2011 -->
1857
1858 <!ELEMENT ORIENTACOES-CONCLUIDAS (ORIENTACOES-CONCLUIDAS-PARA-MESTRADO*,
    ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO*, ORIENTACOES-CONCLUIDAS-PARA-
    POS-DOCTORADO*, OUTRAS-ORIENTACOES-CONCLUIDAS*)>
1859
1860 <!ELEMENT ORIENTACOES-CONCLUIDAS-PARA-MESTRADO (DADOS-BASICOS-DE-
    ORIENTACOES-CONCLUIDAS-PARA-MESTRADO?, DETALHAMENTO-DE-ORIENTACOES-
    CONCLUIDAS-PARA-MESTRADO?, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
    SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1861 <!ATTLIST ORIENTACOES-CONCLUIDAS-PARA-MESTRADO
1862 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1863 >
1864 <!ELEMENT DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO EMPTY>
1865 <!ATTLIST DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO
1866 NATUREZA CDATA #IMPLIED
1867 TIPO (ACADEMICO | PROFISSIONALIZANTE) #IMPLIED
1868 TITULO CDATA #IMPLIED
1869 ANO CDATA #IMPLIED
1870 PAIS CDATA #IMPLIED
1871 IDIOMA CDATA #IMPLIED
1872 HOME-PAGE CDATA #IMPLIED
1873 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1874 DOI CDATA #IMPLIED
1875 TITULO-INGLES CDATA #IMPLIED
1876 >
1877 <!ELEMENT DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO EMPTY>
1878 <!ATTLIST DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-MESTRADO
1879 TIPO-DE-ORIENTACAO (ORIENTADOR_PRINCIPAL | CO_ORIENTADOR) #IMPLIED
1880 NOME-DO-ORIENTADO CDATA #IMPLIED
1881 CODIGO-INSTITUICAO CDATA #IMPLIED
1882 NOME-DA-INSTITUICAO CDATA #IMPLIED
1883 CODIGO-ORGAO CDATA #IMPLIED
1884 NOME-ORGAO CDATA #IMPLIED
1885 CODIGO-CURSO CDATA #IMPLIED
1886 NOME-DO-CURSO CDATA #IMPLIED
1887 FLAG-BOLSA CDATA #IMPLIED
1888 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
1889 NOME-DA-AGENCIA CDATA #IMPLIED
1890 NUMERO-DE-PAGINAS CDATA #IMPLIED
1891 NUMERO-ID-ORIENTADO CDATA #IMPLIED
1892 NOME-DO-CURSO-INGLES CDATA #IMPLIED
1893 >
1894 <!ELEMENT ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO (DADOS-BASICOS-DE-
```

```
        ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO?, DETALHAMENTO-DE-ORIENTACOES-
        CONCLUIDAS-PARA-DOCTORADO?, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
        SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1895 <!ATTLIST ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO
1896 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1897 >
1898 <!ELEMENT DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO EMPTY>
1899 <!ATTLIST DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO
1900 NATUREZA CDATA #IMPLIED
1901 TITULO CDATA #IMPLIED
1902 ANO CDATA #IMPLIED
1903 PAIS CDATA #IMPLIED
1904 IDIOMA CDATA #IMPLIED
1905 HOME-PAGE CDATA #IMPLIED
1906 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1907 DOI CDATA #IMPLIED
1908     TITULO-INGLES CDATA #IMPLIED
1909 >
1910 <!ELEMENT DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO EMPTY>
1911 <!ATTLIST DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-DOCTORADO
1912 TIPO-DE-ORIENTACAO (ORIENTADOR_PRINCIPAL | CO_ORIENTADOR) #IMPLIED
1913 NOME-DO-ORIENTADO CDATA #IMPLIED
1914 CODIGO-INSTITUICAO CDATA #IMPLIED
1915 NOME-DA-INSTITUICAO CDATA #IMPLIED
1916 CODIGO-ORGAO CDATA #IMPLIED
1917 NOME-ORGAO CDATA #IMPLIED
1918 CODIGO-CURSO CDATA #IMPLIED
1919 NOME-DO-CURSO CDATA #IMPLIED
1920 FLAG-BOLSA CDATA #IMPLIED
1921 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
1922 NOME-DA-AGENCIA CDATA #IMPLIED
1923 NUMERO-DE-PAGINAS CDATA #IMPLIED
1924 NUMERO-ID-ORIENTADO CDATA #IMPLIED
1925     NOME-DO-CURSO-INGLES CDATA #IMPLIED
1926 >
1927 <!ELEMENT ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO (DADOS-BASICOS-DE-
        ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO?, DETALHAMENTO-DE-
        ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO?, PALAVRAS-CHAVE?, AREAS-DO-
        -CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1928 <!ATTLIST ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO
1929 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1930 >
1931 <!ELEMENT DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO
        EMPTY>
1932 <!ATTLIST DADOS-BASICOS-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO
1933 NATUREZA CDATA #IMPLIED
1934 TITULO CDATA #IMPLIED
1935 ANO CDATA #IMPLIED
1936 PAIS CDATA #IMPLIED
1937 IDIOMA CDATA #IMPLIED
1938 HOME-PAGE CDATA #IMPLIED
1939 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1940 DOI CDATA #IMPLIED
1941     TITULO-INGLES CDATA #IMPLIED
1942 >
1943 <!ELEMENT DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO
        EMPTY>
1944 <!ATTLIST DETALHAMENTO-DE-ORIENTACOES-CONCLUIDAS-PARA-POS-DOCTORADO
1945 TIPO-DE-ORIENTACAO CDATA #IMPLIED
1946 NOME-DO-ORIENTADO CDATA #IMPLIED
1947 CODIGO-INSTITUICAO CDATA #IMPLIED
1948 NOME-DA-INSTITUICAO CDATA #IMPLIED
1949 CODIGO-ORGAO CDATA #IMPLIED
1950 NOME-ORGAO CDATA #IMPLIED
1951 CODIGO-CURSO CDATA #IMPLIED
1952 NOME-DO-CURSO CDATA #IMPLIED
1953 FLAG-BOLSA CDATA #IMPLIED
1954 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
1955 NOME-DA-AGENCIA CDATA #IMPLIED
1956 NUMERO-DE-PAGINAS CDATA #IMPLIED
1957 NUMERO-ID-ORIENTADO CDATA #IMPLIED
1958     NOME-DO-CURSO-INGLES CDATA #IMPLIED
1959 >
```

```
1960
1961 <!ELEMENT OUTRAS-ORIENTACOES-CONCLUIDAS (DADOS-BASICOS-DE-OUTRAS-
      ORIENTACOES-CONCLUIDAS?, DETALHAMENTO-DE-OUTRAS-ORIENTACOES-
      CONCLUIDAS?, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
      ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1962 <!ATTLIST OUTRAS-ORIENTACOES-CONCLUIDAS
1963 SEQUENCIA-PRODUCAO CDATA #IMPLIED
1964 >
1965 <!ELEMENT DADOS-BASICOS-DE-OUTRAS-ORIENTACOES-CONCLUIDAS EMPTY>
1966 <!ATTLIST DADOS-BASICOS-DE-OUTRAS-ORIENTACOES-CONCLUIDAS
1967 NATUREZA (
      MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_APERFEICOAMENTO_E_ESPECIALIZACAO |
      TRABALHO_DE_CONCLUSAO_DE_CURSO_GRADUACAO | INICIACAO_CIENTIFICA |
      ORIENTACAO-DE-OUTRA-NATUREZA) #IMPLIED
1968 TIPO CDATA #IMPLIED
1969 TITULO CDATA #IMPLIED
1970 ANO CDATA #IMPLIED
1971 PAIS CDATA #IMPLIED
1972 IDIOMA CDATA #IMPLIED
1973 HOME-PAGE CDATA #IMPLIED
1974 FLAG-RELEVANCIA (SIM | NAO) "NAO"
1975 DOI CDATA #IMPLIED
1976 TITULO-INGLES CDATA #IMPLIED
1977 TIPO-INGLES CDATA #IMPLIED
1978 >
1979 <!ELEMENT DETALHAMENTO-DE-OUTRAS-ORIENTACOES-CONCLUIDAS EMPTY>
1980 <!ATTLIST DETALHAMENTO-DE-OUTRAS-ORIENTACOES-CONCLUIDAS
1981 NOME-DO-ORIENTADO CDATA #IMPLIED
1982 CODIGO-INSTITUICAO CDATA #IMPLIED
1983 NOME-DA-INSTITUICAO CDATA #IMPLIED
1984 CODIGO-ORGAO CDATA #IMPLIED
1985 NOME-ORGAO CDATA #IMPLIED
1986 CODIGO-CURSO CDATA #IMPLIED
1987 NOME-DO-CURSO CDATA #IMPLIED
1988 FLAG-BOLSA CDATA #IMPLIED
1989 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
1990 NOME-DA-AGENCIA CDATA #IMPLIED
1991 TIPO-DE-ORIENTACAO-CONCLUIDA CDATA #IMPLIED
1992 NUMERO-DE-PAGINAS CDATA #IMPLIED
1993 NUMERO-ID-ORIENTADO CDATA #IMPLIED
1994 NOME-DO-CURSO-INGLES CDATA #IMPLIED
1995 >
1996
1997 <!ELEMENT DEMAIS-TRABALHOS (DADOS-BASICOS-DE-DEMAIS-TRABALHOS?,
      DETALHAMENTO-DE-DEMAIS-TRABALHOS?, AUTORES*, PALAVRAS-CHAVE?, AREAS-
      DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
1998 <!ATTLIST DEMAIS-TRABALHOS
1999 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2000 >
2001 <!ELEMENT DADOS-BASICOS-DE-DEMAIS-TRABALHOS EMPTY>
2002 <!ATTLIST DADOS-BASICOS-DE-DEMAIS-TRABALHOS
2003 NATUREZA CDATA #IMPLIED
2004 TITULO CDATA #IMPLIED
2005 ANO CDATA #IMPLIED
2006 PAIS CDATA #IMPLIED
2007 IDIOMA CDATA #IMPLIED
2008 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
      FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2009 HOME-PAGE CDATA #IMPLIED
2010 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2011 DOI CDATA #IMPLIED
2012 TITULO-INGLES CDATA #IMPLIED
2013 NATUREZA-INGLES CDATA #IMPLIED
2014 >
2015 <!ELEMENT DETALHAMENTO-DE-DEMAIS-TRABALHOS EMPTY>
2016 <!ATTLIST DETALHAMENTO-DE-DEMAIS-TRABALHOS
2017 FINALIDADE CDATA #IMPLIED
2018 LOCAL CDATA #IMPLIED
2019 >
2020 <!--
      -->
2021 <!--
      -->
```

```
2022 <!-- SEGMENTO DE DADOS-COMPLEMENTARES -->
2023 <!--
-->
2024 <!--
-->
2025 <!ELEMENT DADOS-COMPLEMENTARES (FORMACAO-COMPLEMENTAR*, PARTICIPACAO-EM-
BANCA-TRABALHOS-CONCLUSAO?, PARTICIPACAO-EM-BANCA-JULGADORA?,
PARTICIPACAO-EM-EVENTOS-CONGRESSOS?, ORIENTACOES-EM-ANDAMENTO?,
INFORMACOES-ADICIONAIS-INSTITUICOES?, INFORMACOES-ADICIONAIS-CURSOS?)
>
2026 <!ELEMENT FORMACAO-COMPLEMENTAR (FORMACAO-COMPLEMENTAR-DE-EXTENSAO-
UNIVERSITARIA*, MBA*, FORMACAO-COMPLEMENTAR-CURSO-DE-CURTA-DURACAO*,
OUTROS*)>
2027 <!ELEMENT FORMACAO-COMPLEMENTAR-DE-EXTENSAO-UNIVERSITARIA EMPTY>
2028 <!ATTLIST FORMACAO-COMPLEMENTAR-DE-EXTENSAO-UNIVERSITARIA
2029 SEQUENCIA-FORMACAO CDATA #IMPLIED
2030 NIVEL CDATA #IMPLIED
2031 CARGA-HORARIA CDATA #IMPLIED
2032 CODIGO-INSTITUICAO CDATA #IMPLIED
2033 NOME-INSTITUICAO CDATA #IMPLIED
2034 CODIGO-ORGAO CDATA #IMPLIED
2035 NOME-ORGAO CDATA #IMPLIED
2036 CODIGO-CURSO CDATA #IMPLIED
2037 NOME-CURSO CDATA #IMPLIED
2038 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
2039 ANO-DE-INICIO CDATA #IMPLIED
2040 ANO-DE-CONCLUSAO CDATA #IMPLIED
2041 NOME-CURSO-INGLES CDATA #IMPLIED
2042 >
2043 <!ELEMENT MBA (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
ATIVIDADE?)>
2044 <!ATTLIST MBA
2045 SEQUENCIA-FORMACAO CDATA #IMPLIED
2046 NIVEL CDATA #IMPLIED
2047 CARGA-HORARIA CDATA #IMPLIED
2048 CODIGO-INSTITUICAO CDATA #IMPLIED
2049 NOME-INSTITUICAO CDATA #IMPLIED
2050 CODIGO-ORGAO CDATA #IMPLIED
2051 NOME-ORGAO CDATA #IMPLIED
2052 CODIGO-CURSO CDATA #IMPLIED
2053 NOME-CURSO CDATA #IMPLIED
2054 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
2055 ANO-DE-INICIO CDATA #IMPLIED
2056 ANO-DE-CONCLUSAO CDATA #IMPLIED
2057 ANO-DE-OBTENCAO-DO-TITULO CDATA #IMPLIED
2058 FLAG-BOLSA CDATA #IMPLIED
2059 CODIGO-AGENCIA-FINANCIADORA CDATA #IMPLIED
2060 NOME-AGENCIA CDATA #IMPLIED
2061 TITULO-DA-MONOGRRAFIA CDATA #IMPLIED
2062 NOME-COMPLETO-DO-ORIENTADOR CDATA #IMPLIED
2063 NOME-CURSO-INGLES CDATA #IMPLIED
2064 >
2065 <!ELEMENT FORMACAO-COMPLEMENTAR-CURSO-DE-CURTA-DURACAO EMPTY>
2066 <!ATTLIST FORMACAO-COMPLEMENTAR-CURSO-DE-CURTA-DURACAO
2067 SEQUENCIA-FORMACAO CDATA #IMPLIED
2068 NIVEL CDATA #IMPLIED
2069 CARGA-HORARIA CDATA #IMPLIED
2070 CODIGO-INSTITUICAO CDATA #IMPLIED
2071 NOME-INSTITUICAO CDATA #IMPLIED
2072 CODIGO-ORGAO CDATA #IMPLIED
2073 NOME-ORGAO CDATA #IMPLIED
2074 CODIGO-CURSO CDATA #IMPLIED
2075 NOME-CURSO CDATA #IMPLIED
2076 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
2077 ANO-DE-INICIO CDATA #IMPLIED
2078 ANO-DE-CONCLUSAO CDATA #IMPLIED
2079 NOME-CURSO-INGLES CDATA #IMPLIED
2080 >
2081 <!ELEMENT OUTROS EMPTY>
2082 <!ATTLIST OUTROS
2083 SEQUENCIA-FORMACAO CDATA #IMPLIED
2084 NIVEL CDATA #IMPLIED
2085 CARGA-HORARIA CDATA #IMPLIED
```

```
2086 CODIGO-INSTITUICAO CDATA #IMPLIED
2087 NOME-INSTITUICAO CDATA #IMPLIED
2088 CODIGO-ORGAO CDATA #IMPLIED
2089 NOME-ORGAO CDATA #IMPLIED
2090 CODIGO-CURSO CDATA #IMPLIED
2091 NOME-CURSO CDATA #IMPLIED
2092 STATUS-DO-CURSO (EM_ANDAMENTO | CONCLUIDO | INCOMPLETO) #IMPLIED
2093 ANO-DE-INICIO CDATA #IMPLIED
2094 ANO-DE-CONCLUSAO CDATA #IMPLIED
2095 NOME-CURSO-INGLES CDATA #IMPLIED
2096 >
2097 <!ELEMENT PARTICIPACAO-EM-BANCA-TRABALHOS-CONCLUSAO (PARTICIPACAO-EM-
BANCA-DE-MESTRADO*, PARTICIPACAO-EM-BANCA-DE-DOCTORADO*, PARTICIPACAO
-EM-BANCA-DE-EXAME-QUALIFICACAO*, PARTICIPACAO-EM-BANCA-DE-
APERFEICOAMENTO-ESPECIALIZACAO*, PARTICIPACAO-EM-BANCA-DE-GRADUACAO*,
OUTRAS-PARTICIPACOES-EM-BANCA*)>
2098 <!ELEMENT PARTICIPACAO-EM-BANCA-DE-MESTRADO (DADOS-BASICOS-DA-
PARTICIPACAO-EM-BANCA-DE-MESTRADO?, DETALHAMENTO-DA-PARTICIPACAO-EM-
BANCA-DE-MESTRADO?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2099 <!ATTLIST PARTICIPACAO-EM-BANCA-DE-MESTRADO
2100 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2101 >
2102 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-MESTRADO EMPTY>
2103 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-MESTRADO
2104 NATUREZA CDATA #IMPLIED
2105 TIPO (ACADEMICO | PROFISSIONALIZANTE | NAO_INFORMADO) #IMPLIED
2106 TITULO CDATA #IMPLIED
2107 ANO CDATA #IMPLIED
2108 PAIS CDATA #IMPLIED
2109 IDIOMA CDATA #IMPLIED
2110 HOME-PAGE CDATA #IMPLIED
2111 DOI CDATA #IMPLIED
2112 TITULO-INGLES CDATA #IMPLIED
2113 >
2114 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-MESTRADO EMPTY>
2115 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-MESTRADO
2116 NOME-DO-CANDIDATO CDATA #IMPLIED
2117 CODIGO-INSTITUICAO CDATA #IMPLIED
2118 NOME-INSTITUICAO CDATA #IMPLIED
2119 CODIGO-ORGAO CDATA #IMPLIED
2120 NOME-ORGAO CDATA #IMPLIED
2121 CODIGO-CURSO CDATA #IMPLIED
2122 NOME-CURSO CDATA #IMPLIED
2123 NOME-CURSO-INGLES CDATA #IMPLIED
2124 >
2125 <!ELEMENT PARTICIPANTE-BANCA EMPTY>
2126 <!ATTLIST PARTICIPANTE-BANCA
2127 NOME-COMPLETO-DO-PARTICIPANTE-DA-BANCA CDATA #IMPLIED
2128 NOME-PARA-CITACAO-DO-PARTICIPANTE-DA-BANCA CDATA #IMPLIED
2129 ORDEM-PARTICIPANTE CDATA #IMPLIED
2130 CPF-DO-PARTICIPANTE-DA-BANCA CDATA #IMPLIED
2131 NRO-ID-CNPQ CDATA #IMPLIED
2132 >
2133 <!ELEMENT PARTICIPACAO-EM-BANCA-DE-DOCTORADO (DADOS-BASICOS-DA-
PARTICIPACAO-EM-BANCA-DE-DOCTORADO?, DETALHAMENTO-DA-PARTICIPACAO-EM-
BANCA-DE-DOCTORADO?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2134 <!ATTLIST PARTICIPACAO-EM-BANCA-DE-DOCTORADO
2135 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2136 >
2137 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-DOCTORADO EMPTY>
2138 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-DOCTORADO
2139 NATUREZA CDATA #IMPLIED
2140 TITULO CDATA #IMPLIED
2141 ANO CDATA #IMPLIED
2142 PAIS CDATA #IMPLIED
2143 IDIOMA CDATA #IMPLIED
2144 HOME-PAGE CDATA #IMPLIED
2145 DOI CDATA #IMPLIED
2146 TITULO-INGLES CDATA #IMPLIED
2147 >
2148 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-DOCTORADO EMPTY>
```

```
2149 <! ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-DOCTORADO
2150 NOME-DO-CANDIDATO CDATA #IMPLIED
2151 CODIGO-INSTITUICAO CDATA #IMPLIED
2152 NOME-INSTITUICAO CDATA #IMPLIED
2153 CODIGO-ORGAO CDATA #IMPLIED
2154 NOME-ORGAO CDATA #IMPLIED
2155 CODIGO-CURSO CDATA #IMPLIED
2156 NOME-CURSO CDATA #IMPLIED
2157 NOME-CURSO-INGLES CDATA #IMPLIED
2158 >
2159 <! ELEMENT PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO (DADOS-BASICOS-DA-
PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO?, DETALHAMENTO-DA-
PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO?, PARTICIPANTE-BANCA*,
PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?,
INFORMACOES-ADICIONAIS?)>
2160 <! ATTLIST PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO
2161 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2162 >
2163 <! ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO
EMPTY>
2164 <! ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO
2165 NATUREZA CDATA #IMPLIED
2166 TITULO CDATA #IMPLIED
2167 ANO CDATA #IMPLIED
2168 PAIS CDATA #IMPLIED
2169 IDIOMA CDATA #IMPLIED
2170 HOME-PAGE CDATA #IMPLIED
2171 DOI CDATA #IMPLIED
2172 TITULO-INGLES CDATA #IMPLIED
2173 >
2174 <! ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO
EMPTY>
2175 <! ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-EXAME-QUALIFICACAO
2176 NOME-DO-CANDIDATO CDATA #IMPLIED
2177 CODIGO-INSTITUICAO CDATA #IMPLIED
2178 NOME-INSTITUICAO CDATA #IMPLIED
2179 CODIGO-ORGAO CDATA #IMPLIED
2180 NOME-ORGAO CDATA #IMPLIED
2181 CODIGO-CURSO CDATA #IMPLIED
2182 NOME-CURSO CDATA #IMPLIED
2183 NOME-CURSO-INGLES CDATA #IMPLIED
2184 >
2185 <! ELEMENT PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-ESPECIALIZACAO (DADOS
-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-ESPECIALIZACAO?,
DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-
ESPECIALIZACAO?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2186 <! ATTLIST PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-ESPECIALIZACAO
2187 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2188 >
2189 <! ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-
ESPECIALIZACAO EMPTY>
2190 <! ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-
ESPECIALIZACAO
2191 NATUREZA CDATA #IMPLIED
2192 TITULO CDATA #IMPLIED
2193 ANO CDATA #IMPLIED
2194 PAIS CDATA #IMPLIED
2195 IDIOMA CDATA #IMPLIED
2196 HOME-PAGE CDATA #IMPLIED
2197 DOI CDATA #IMPLIED
2198 TITULO-INGLES CDATA #IMPLIED
2199 >
2200 <! ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-
ESPECIALIZACAO EMPTY>
2201 <! ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-APERFEICOAMENTO-
ESPECIALIZACAO
2202 NOME-DO-CANDIDATO CDATA #IMPLIED
2203 CODIGO-INSTITUICAO CDATA #IMPLIED
2204 NOME-INSTITUICAO CDATA #IMPLIED
2205 CODIGO-ORGAO CDATA #IMPLIED
2206 NOME-ORGAO CDATA #IMPLIED
2207 CODIGO-CURSO CDATA #IMPLIED
```

```
2208 NOME-CURSO CDATA #IMPLIED
2209 NOME-CURSO-INGLES CDATA #IMPLIED
2210 >
2211 <!ELEMENT PARTICIPACAO-EM-BANCA-DE-GRADUACAO (DADOS-BASICOS-DA-
PARTICIPACAO-EM-BANCA-DE-GRADUACAO?, DETALHAMENTO-DA-PARTICIPACAO-EM-
BANCA-DE-GRADUACAO?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2212 <!ATTLIST PARTICIPACAO-EM-BANCA-DE-GRADUACAO
2213 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2214 >
2215 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-GRADUACAO EMPTY>
2216 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-BANCA-DE-GRADUACAO
2217 NATUREZA CDATA #IMPLIED
2218 TITULO CDATA #IMPLIED
2219 ANO CDATA #IMPLIED
2220 PAIS CDATA #IMPLIED
2221 IDIOMA CDATA #IMPLIED
2222 HOME-PAGE CDATA #IMPLIED
2223 DOI CDATA #IMPLIED
2224 TITULO-INGLES CDATA #IMPLIED
2225 >
2226 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-GRADUACAO EMPTY>
2227 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-BANCA-DE-GRADUACAO
2228 NOME-DO-CANDIDATO CDATA #IMPLIED
2229 CODIGO-INSTITUICAO CDATA #IMPLIED
2230 NOME-INSTITUICAO CDATA #IMPLIED
2231 CODIGO-ORGAO CDATA #IMPLIED
2232 NOME-ORGAO CDATA #IMPLIED
2233 CODIGO-CURSO CDATA #IMPLIED
2234 NOME-CURSO CDATA #IMPLIED
2235 NOME-CURSO-INGLES CDATA #IMPLIED
2236 >
2237 <!ELEMENT OUTRAS-PARTICIPACOES-EM-BANCA (DADOS-BASICOS-DE-OUTRAS-
PARTICIPACOES-EM-BANCA?, DETALHAMENTO-DE-OUTRAS-PARTICIPACOES-EM-
BANCA?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2238 <!ATTLIST OUTRAS-PARTICIPACOES-EM-BANCA
2239 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2240 >
2241 <!ELEMENT DADOS-BASICOS-DE-OUTRAS-PARTICIPACOES-EM-BANCA EMPTY>
2242 <!ATTLIST DADOS-BASICOS-DE-OUTRAS-PARTICIPACOES-EM-BANCA
2243 NATUREZA CDATA #IMPLIED
2244 TIPO CDATA #IMPLIED
2245 TITULO CDATA #IMPLIED
2246 ANO CDATA #IMPLIED
2247 PAIS CDATA #IMPLIED
2248 IDIOMA CDATA #IMPLIED
2249 HOME-PAGE CDATA #IMPLIED
2250 DOI CDATA #IMPLIED
2251 TITULO-INGLES CDATA #IMPLIED
2252 TIPO-INGLES CDATA #IMPLIED
2253 >
2254 <!ELEMENT DETALHAMENTO-DE-OUTRAS-PARTICIPACOES-EM-BANCA EMPTY>
2255 <!ATTLIST DETALHAMENTO-DE-OUTRAS-PARTICIPACOES-EM-BANCA
2256 NOME-DO-CANDIDATO CDATA #IMPLIED
2257 CODIGO-INSTITUICAO CDATA #IMPLIED
2258 NOME-INSTITUICAO CDATA #IMPLIED
2259 CODIGO-ORGAO CDATA #IMPLIED
2260 NOME-ORGAO CDATA #IMPLIED
2261 CODIGO-CURSO CDATA #IMPLIED
2262 NOME-CURSO CDATA #IMPLIED
2263 NOME-CURSO-INGLES CDATA #IMPLIED
2264 >
2265 <!ELEMENT PARTICIPACAO-EM-BANCA-JULGADORA (BANCA-JULGADORA-PARA-
PROFESSOR-TITULAR*, BANCA-JULGADORA-PARA-CONCURSO-PUBLICO*, BANCA-
JULGADORA-PARA-LIVRE-DOCENCIA*, BANCA-JULGADORA-PARA-AVALIACAO-CURSOS
*, OUTRAS-BANCAS-JULGADORAS*)>
2266 <!ELEMENT BANCA-JULGADORA-PARA-PROFESSOR-TITULAR (DADOS-BASICOS-DA-BANCA-
JULGADORA-PARA-PROFESSOR-TITULAR?, DETALHAMENTO-DA-BANCA-JULGADORA-
PARA-PROFESSOR-TITULAR?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-
DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2267 <!ATTLIST BANCA-JULGADORA-PARA-PROFESSOR-TITULAR
2268 SEQUENCIA-PRODUCAO CDATA #IMPLIED
```

```
2269 >
2270 <!ELEMENT DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-PROFESSOR-TITULAR EMPTY>
2271 <!ATTLIST DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-PROFESSOR-TITULAR
2272 NATUREZA CDATA #IMPLIED
2273 TITULO CDATA #IMPLIED
2274 ANO CDATA #IMPLIED
2275 PAIS CDATA #IMPLIED
2276 IDIOMA CDATA #IMPLIED
2277 HOME-PAGE CDATA #IMPLIED
2278 DOI CDATA #IMPLIED
2279 TITULO-INGLES CDATA #IMPLIED
2280 >
2281 <!ELEMENT DETALHAMENTO-DA-BANCA-JULGADORA-PARA-PROFESSOR-TITULAR EMPTY>
2282 <!ATTLIST DETALHAMENTO-DA-BANCA-JULGADORA-PARA-PROFESSOR-TITULAR
2283 CODIGO-INSTITUICAO CDATA #IMPLIED
2284 NOME-INSTITUICAO CDATA #IMPLIED
2285 >
2286 <!ELEMENT BANCA-JULGADORA-PARA-CONCURSO-PUBLICO (DADOS-BASICOS-DA-BANCA-
JULGADORA-PARA-CONCURSO-PUBLICO?, DETALHAMENTO-DA-BANCA-JULGADORA-
PARA-CONCURSO-PUBLICO?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-
DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2287 <!ATTLIST BANCA-JULGADORA-PARA-CONCURSO-PUBLICO
2288 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2289 >
2290 <!ELEMENT DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-CONCURSO-PUBLICO EMPTY>
2291 <!ATTLIST DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-CONCURSO-PUBLICO
2292 NATUREZA CDATA #IMPLIED
2293 TITULO CDATA #IMPLIED
2294 ANO CDATA #IMPLIED
2295 PAIS CDATA #IMPLIED
2296 IDIOMA CDATA #IMPLIED
2297 HOME-PAGE CDATA #IMPLIED
2298 DOI CDATA #IMPLIED
2299 TITULO-INGLES CDATA #IMPLIED
2300 >
2301 <!ELEMENT DETALHAMENTO-DA-BANCA-JULGADORA-PARA-CONCURSO-PUBLICO EMPTY>
2302 <!ATTLIST DETALHAMENTO-DA-BANCA-JULGADORA-PARA-CONCURSO-PUBLICO
2303 CODIGO-INSTITUICAO CDATA #IMPLIED
2304 NOME-INSTITUICAO CDATA #IMPLIED
2305 >
2306 <!ELEMENT BANCA-JULGADORA-PARA-LIVRE-DOCENCIA (DADOS-BASICOS-DA-BANCA-
JULGADORA-PARA-LIVRE-DOCENCIA?, DETALHAMENTO-DA-BANCA-JULGADORA-PARA-
LIVRE-DOCENCIA?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-DO-
CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2307 <!ATTLIST BANCA-JULGADORA-PARA-LIVRE-DOCENCIA
2308 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2309 >
2310 <!ELEMENT DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-LIVRE-DOCENCIA EMPTY>
2311 <!ATTLIST DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-LIVRE-DOCENCIA
2312 NATUREZA CDATA #IMPLIED
2313 TITULO CDATA #IMPLIED
2314 ANO CDATA #IMPLIED
2315 PAIS CDATA #IMPLIED
2316 IDIOMA CDATA #IMPLIED
2317 HOME-PAGE CDATA #IMPLIED
2318 DOI CDATA #IMPLIED
2319 TITULO-INGLES CDATA #IMPLIED
2320 >
2321 <!ELEMENT DETALHAMENTO-DA-BANCA-JULGADORA-PARA-LIVRE-DOCENCIA EMPTY>
2322 <!ATTLIST DETALHAMENTO-DA-BANCA-JULGADORA-PARA-LIVRE-DOCENCIA
2323 CODIGO-INSTITUICAO CDATA #IMPLIED
2324 NOME-INSTITUICAO CDATA #IMPLIED
2325 >
2326 <!ELEMENT BANCA-JULGADORA-PARA-AVALIACAO-CURSOS (DADOS-BASICOS-DA-BANCA-
JULGADORA-PARA-AVALIACAO-CURSOS?, DETALHAMENTO-DA-BANCA-JULGADORA-
PARA-AVALIACAO-CURSOS?, PARTICIPANTE-BANCA*, PALAVRAS-CHAVE?, AREAS-
DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2327 <!ATTLIST BANCA-JULGADORA-PARA-AVALIACAO-CURSOS
2328 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2329 >
2330 <!ELEMENT DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-AVALIACAO-CURSOS EMPTY>
2331 <!ATTLIST DADOS-BASICOS-DA-BANCA-JULGADORA-PARA-AVALIACAO-CURSOS
2332 NATUREZA CDATA #IMPLIED
```

```
2333 TITULO CDATA #IMPLIED
2334 ANO CDATA #IMPLIED
2335 PAIS CDATA #IMPLIED
2336 IDIOMA CDATA #IMPLIED
2337 HOME-PAGE CDATA #IMPLIED
2338 DOI CDATA #IMPLIED
2339 TITULO-INGLES CDATA #IMPLIED
2340 >
2341 <!ELEMENT DETALHAMENTO-DA-BANCA-JULGADORA-PARA-AVALIACAO-CURSOS EMPTY>
2342 <!ATTLIST DETALHAMENTO-DA-BANCA-JULGADORA-PARA-AVALIACAO-CURSOS
2343 CODIGO-INSTITUICAO CDATA #IMPLIED
2344 NOME-INSTITUICAO CDATA #IMPLIED
2345 >
2346 <!ELEMENT OUTRAS-BANCAS-JULGADORAS (DADOS-BASICOS-DE-OUTRAS-BANCAS-
JULGADORAS?, DETALHAMENTO-DE-OUTRAS-BANCAS-JULGADORAS?, PARTICIPANTE-
BANCA*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-ATIVIDADE
?, INFORMACOES-ADICIONAIS?)>
2347 <!ATTLIST OUTRAS-BANCAS-JULGADORAS
2348 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2349 >
2350 <!ELEMENT DADOS-BASICOS-DE-OUTRAS-BANCAS-JULGADORAS EMPTY>
2351 <!ATTLIST DADOS-BASICOS-DE-OUTRAS-BANCAS-JULGADORAS
2352 TIPO CDATA #IMPLIED
2353 NATUREZA CDATA #IMPLIED
2354 TITULO CDATA #IMPLIED
2355 ANO CDATA #IMPLIED
2356 PAIS CDATA #IMPLIED
2357 IDIOMA CDATA #IMPLIED
2358 HOME-PAGE CDATA #IMPLIED
2359 DOI CDATA #IMPLIED
2360 TITULO-INGLES CDATA #IMPLIED
2361 TIPO-INGLES CDATA #IMPLIED
2362 >
2363 <!ELEMENT DETALHAMENTO-DE-OUTRAS-BANCAS-JULGADORAS EMPTY>
2364 <!ATTLIST DETALHAMENTO-DE-OUTRAS-BANCAS-JULGADORAS
2365 CODIGO-INSTITUICAO CDATA #IMPLIED
2366 NOME-INSTITUICAO CDATA #IMPLIED
2367 >
2368 <!ELEMENT PARTICIPACAO-EM-EVENTOS-CONGRESSOS (PARTICIPACAO-EM-CONGRESSO
*, PARTICIPACAO-EM-FEIRA*, PARTICIPACAO-EM-SEMINARIO*, PARTICIPACAO-EM-
SIMPOSIO*, PARTICIPACAO-EM-OFICINA*, PARTICIPACAO-EM-ENCONTRO*,
PARTICIPACAO-EM-EXPOSICAO*, PARTICIPACAO-EM-OLIMPIADA*, OUTRAS-
PARTICIPACOES-EM-EVENTOS-CONGRESSOS*)>
2369 <!ELEMENT PARTICIPANTE-DE-EVENTOS-CONGRESSOS EMPTY>
2370 <!ATTLIST PARTICIPANTE-DE-EVENTOS-CONGRESSOS
2371 NOME-COMPLETO-DO-PARTICIPANTE-DE-EVENTOS-CONGRESSOS CDATA #IMPLIED
2372 NOME-PARA-CITACAO-DO-PARTICIPANTE-DE-EVENTOS-CONGRESSOS CDATA #IMPLIED
2373 ORDEM-PARTICIPANTE CDATA #IMPLIED
2374 CPF-DO-PARTICIPANTE-PARTICIPANTE-DE-EVENTOS-CONGRESSOS CDATA #IMPLIED
2375 NRO-ID-CNPQ CDATA #IMPLIED
2376 >
2377 <!ELEMENT PARTICIPACAO-EM-CONGRESSO (DADOS-BASICOS-DA-PARTICIPACAO-EM-
CONGRESSO?, DETALHAMENTO-DA-PARTICIPACAO-EM-CONGRESSO?, PARTICIPANTE-
DE-EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2378 <!ATTLIST PARTICIPACAO-EM-CONGRESSO
2379 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2380 >
2381 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-CONGRESSO EMPTY>
2382 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-CONGRESSO
2383 NATUREZA CDATA #IMPLIED
2384 TITULO CDATA #IMPLIED
2385 ANO CDATA #IMPLIED
2386 PAIS CDATA #IMPLIED
2387 IDIOMA CDATA #IMPLIED
2388 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2389 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2390 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2391 TIPO-PARTICIPACAO CDATA #IMPLIED
2392 FORMA-PARTICIPACAO CDATA #IMPLIED
2393 DOI CDATA #IMPLIED
2394 TITULO-INGLES CDATA #IMPLIED
```

```
2395 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2396 >
2397 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-CONGRESSO EMPTY>
2398 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-CONGRESSO
2399 NOME-DO-EVENTO CDATA #IMPLIED
2400 CODIGO-INSTITUICAO CDATA #IMPLIED
2401 NOME-INSTITUICAO CDATA #IMPLIED
2402 LOCAL-DO-EVENTO CDATA #IMPLIED
2403 CIDADE-DO-EVENTO CDATA #IMPLIED
2404 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2405 >
2406 <!ELEMENT PARTICIPACAO-EM-FEIRA (DADOS-BASICOS-DA-PARTICIPACAO-EM-FEIRA
?, DETALHAMENTO-DA-PARTICIPACAO-EM-FEIRA?, PARTICIPANTE-DE-EVENTOS-
CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES-DE-
ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2407 <!ATTLIST PARTICIPACAO-EM-FEIRA
2408 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2409 >
2410 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-FEIRA EMPTY>
2411 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-FEIRA
2412 NATUREZA CDATA #IMPLIED
2413 TITULO CDATA #IMPLIED
2414 ANO CDATA #IMPLIED
2415 PAIS CDATA #IMPLIED
2416 IDIOMA CDATA #IMPLIED
2417 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2418 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2419 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2420 TIPO-PARTICIPACAO CDATA #IMPLIED
2421 FORMA-PARTICIPACAO CDATA #IMPLIED
2422 DOI CDATA #IMPLIED
2423 TITULO-INGLES CDATA #IMPLIED
2424 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2425 >
2426 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-FEIRA EMPTY>
2427 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-FEIRA
2428 NOME-DO-EVENTO CDATA #IMPLIED
2429 CODIGO-INSTITUICAO CDATA #IMPLIED
2430 NOME-INSTITUICAO CDATA #IMPLIED
2431 LOCAL-DO-EVENTO CDATA #IMPLIED
2432 CIDADE-DO-EVENTO CDATA #IMPLIED
2433 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2434 >
2435 <!ELEMENT PARTICIPACAO-EM-SEMINARIO (DADOS-BASICOS-DA-PARTICIPACAO-EM-
SEMINARIO?, DETALHAMENTO-DA-PARTICIPACAO-EM-SEMINARIO?, PARTICIPANTE-
DE-EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2436 <!ATTLIST PARTICIPACAO-EM-SEMINARIO
2437 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2438 >
2439 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-SEMINARIO EMPTY>
2440 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-SEMINARIO
2441 NATUREZA CDATA #IMPLIED
2442 TITULO CDATA #IMPLIED
2443 ANO CDATA #IMPLIED
2444 PAIS CDATA #IMPLIED
2445 IDIOMA CDATA #IMPLIED
2446 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2447 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2448 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2449 TIPO-PARTICIPACAO CDATA #IMPLIED
2450 FORMA-PARTICIPACAO CDATA #IMPLIED
2451 DOI CDATA #IMPLIED
2452 TITULO-INGLES CDATA #IMPLIED
2453 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2454 >
2455 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-SEMINARIO EMPTY>
2456 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-SEMINARIO
2457 NOME-DO-EVENTO CDATA #IMPLIED
2458 CODIGO-INSTITUICAO CDATA #IMPLIED
```

```
2459 NOME-INSTITUICAO CDATA #IMPLIED
2460 LOCAL-DO-EVENTO CDATA #IMPLIED
2461 CIDADE-DO-EVENTO CDATA #IMPLIED
2462 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2463 >
2464 <!ELEMENT PARTICIPACAO-EM-SIMPOSIO (DADOS-BASICOS-DA-PARTICIPACAO-EM-
SIMPOSIO?, DETALHAMENTO-DA-PARTICIPACAO-EM-SIMPOSIO?, PARTICIPANTE-DE
-EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2465 <!ATTLIST PARTICIPACAO-EM-SIMPOSIO
2466 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2467 >
2468 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-SIMPOSIO EMPTY>
2469 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-SIMPOSIO
2470 NATUREZA CDATA #IMPLIED
2471 TITULO CDATA #IMPLIED
2472 ANO CDATA #IMPLIED
2473 PAIS CDATA #IMPLIED
2474 IDIOMA CDATA #IMPLIED
2475 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2476 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2477 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2478 TIPO-PARTICIPACAO CDATA #IMPLIED
2479 FORMA-PARTICIPACAO CDATA #IMPLIED
2480 DOI CDATA #IMPLIED
2481 TITULO-INGLES CDATA #IMPLIED
2482 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2483 >
2484 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-SIMPOSIO EMPTY>
2485 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-SIMPOSIO
2486 NOME-DO-EVENTO CDATA #IMPLIED
2487 CODIGO-INSTITUICAO CDATA #IMPLIED
2488 NOME-INSTITUICAO CDATA #IMPLIED
2489 LOCAL-DO-EVENTO CDATA #IMPLIED
2490 CIDADE-DO-EVENTO CDATA #IMPLIED
2491 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2492 >
2493 <!ELEMENT PARTICIPACAO-EM-OFICINA (DADOS-BASICOS-DA-PARTICIPACAO-EM-
OFICINA?, DETALHAMENTO-DA-PARTICIPACAO-EM-OFICINA?, PARTICIPANTE-DE-
EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?, SETORES
-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2494 <!ATTLIST PARTICIPACAO-EM-OFICINA
2495 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2496 >
2497 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-OFICINA EMPTY>
2498 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-OFICINA
2499 NATUREZA CDATA #IMPLIED
2500 TITULO CDATA #IMPLIED
2501 ANO CDATA #IMPLIED
2502 PAIS CDATA #IMPLIED
2503 IDIOMA CDATA #IMPLIED
2504 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2505 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2506 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2507 TIPO-PARTICIPACAO CDATA #IMPLIED
2508 FORMA-PARTICIPACAO CDATA #IMPLIED
2509 DOI CDATA #IMPLIED
2510 TITULO-INGLES CDATA #IMPLIED
2511 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2512 >
2513 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-OFICINA EMPTY>
2514 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-OFICINA
2515 NOME-DO-EVENTO CDATA #IMPLIED
2516 CODIGO-INSTITUICAO CDATA #IMPLIED
2517 NOME-INSTITUICAO CDATA #IMPLIED
2518 LOCAL-DO-EVENTO CDATA #IMPLIED
2519 CIDADE-DO-EVENTO CDATA #IMPLIED
2520 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2521 >
2522 <!ELEMENT PARTICIPACAO-EM-ENCONTRO (DADOS-BASICOS-DA-PARTICIPACAO-EM-
ENCONTRO?, DETALHAMENTO-DA-PARTICIPACAO-EM-ENCONTRO?, PARTICIPANTE-DE
```

```
        -EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
        SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2523 <!ATTLIST PARTICIPACAO-EM-ENCONTRO
2524 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2525 >
2526 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-ENCONTRO EMPTY>
2527 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-ENCONTRO
2528 NATUREZA CDATA #IMPLIED
2529 TITULO CDATA #IMPLIED
2530 ANO CDATA #IMPLIED
2531 PAIS CDATA #IMPLIED
2532 IDIOMA CDATA #IMPLIED
2533 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
        FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2534 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2535 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2536 TIPO-PARTICIPACAO CDATA #IMPLIED
2537 FORMA-PARTICIPACAO CDATA #IMPLIED
2538 DOI CDATA #IMPLIED
2539 TITULO-INGLES CDATA #IMPLIED
2540 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2541 >
2542 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-ENCONTRO EMPTY>
2543 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-ENCONTRO
2544 NOME-DO-EVENTO CDATA #IMPLIED
2545 CODIGO-INSTITUICAO CDATA #IMPLIED
2546 NOME-INSTITUICAO CDATA #IMPLIED
2547 LOCAL-DO-EVENTO CDATA #IMPLIED
2548 CIDADE-DO-EVENTO CDATA #IMPLIED
2549 NOME-DO-EVENTO-INGLES CDATA #IMPLIED
2550 >
2551 <!ELEMENT PARTICIPACAO-EM-EXPOSICAO (DADOS-BASICOS-DA-PARTICIPACAO-EM-
        EXPOSICAO?, DETALHAMENTO-DA-PARTICIPACAO-EM-EXPOSICAO?, PARTICIPANTE-
        DE-EVENTOS-CONGRESSOS*, PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO?,
        SETORES-DE-ATIVIDADE?, INFORMACOES-ADICIONAIS?)>
2552 <!ATTLIST PARTICIPACAO-EM-EXPOSICAO
2553 SEQUENCIA-PRODUCAO CDATA #IMPLIED
2554 >
2555 <!ELEMENT DADOS-BASICOS-DA-PARTICIPACAO-EM-EXPOSICAO EMPTY>
2556 <!ATTLIST DADOS-BASICOS-DA-PARTICIPACAO-EM-EXPOSICAO
2557 NATUREZA CDATA #IMPLIED
2558 TITULO CDATA #IMPLIED
2559 ANO CDATA #IMPLIED
2560 PAIS CDATA #IMPLIED
2561 IDIOMA CDATA #IMPLIED
2562 MEIO-DE-DIVULGACAO (IMPRESSO | WEB | MEIO_MAGNETICO | MEIO_DIGITAL |
        FILME | HIPERTEXTO | OUTRO | VARIOS | NAO_INFORMADO) "NAO_INFORMADO"
2563 HOME-PAGE-DO-TRABALHO CDATA #IMPLIED
2564 FLAG-RELEVANCIA (SIM | NAO) "NAO"
2565 TIPO-PARTICIPACAO CDATA #IMPLIED
2566 FORMA-PARTICIPACAO CDATA #IMPLIED
2567 DOI CDATA #IMPLIED
2568 TITULO-INGLES CDATA #IMPLIED
2569 FLAG-DIVULGACAO-CIENTIFICA (SIM | NAO) "NAO"
2570 >
2571 <!ELEMENT DETALHAMENTO-DA-PARTICIPACAO-EM-EXPOSICAO EMPTY>
2572 <!ATTLIST DETALHAMENTO-DA-PARTICIPACAO-EM-EXPOSICAO
2573 NOME-DO-EVENTO CDATA #IMPLIED
2574 CODIGO-INSTITUICAO CDATA #
```
