UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Laura Lima Dias

Análise de Abordagens Automáticas de Anotação Semântica para Textos Ruidosos e seus Impactos na Similaridade entre Vídeos

Juiz de Fora

2017

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Laura Lima Dias

Análise de Abordagens Automáticas de Anotação Semântica para Textos Ruidosos e seus Impactos na Similaridade entre Vídeos

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Eduardo Barrére

Coorientador: Jairo Francisco de Souza

Juiz de Fora

2017

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Dias, Laura Lima.
Análise de Abordagens Automáticas de Anotação Semântica para Textos Ruidosos e seus Impactos na Similaridade entre Vídeos / Laura Lima Dias. -- 2017.
74 p. : il.

Orientador: Eduardo Barrére
Coorientador: Jairo Francisco de Souza
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós Graduação em Ciência da Computação, 2017.

1. Recuperação de Informação. 2. Repositório de Videos. 3. Categorização de Texto Ruidoso. 4. Processamento de Linguagem Natural. I. Barrére, Eduardo, orient. II. Souza, Jairo Francisco de, coorient. III. Título.

Laura Lima Dias

Análise de Abordagens Automáticas de Anotação Semântica para Textos Ruidosos e seus Impactos na Similaridade entre Vídeos

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 31 de Agosto de 2017.

BANCA EXAMINADORA

Prof. D.Sc. Eduardo Barrére - Orientador
Universidade Federal de Juiz de Fora

Prof. D.Sc. Jairo Francisco de Souza - Coorientador
Universidade Federal de Juiz de Fora

Prof. D.Sc. Marcelo Ferreira Moreno
Universidade Federal de Juiz de Fora

Prof. D.Sc. Celso Alberto Saibel Santos
Universidade Federal do Espírito Santo

Dedico esse trabalho aos meus amados pais, por estarem ao meu lado, apoiando e incentivando e a toda minha família e amigos, pelo carinho e apoio.

AGRADECIMENTOS

É chegada a reta final de mais uma etapa de minha vida, para a qual se deve muito esforço e dedicação. Primeiramente, agradeço a Deus pelo dom da vida, por toda sabedoria e conhecimento e a Virgem Maria por interceder junto a Deus por mim.

Agradeço aos meus pais José Adilson e Maria das Dôres pelo amor, bons exemplos, por todo apoio, incentivo, dedicação exclusiva e pelos valiosos ensinamentos. Amo muito vocês!

A minha amada Avó que não pode presenciar essa conquista, mas que sempre quis o melhor para mim.

Agradeço aos meus tios José Braz e Solange e a minha prima Thayná por permitirem que a casa de vocês fosse também a minha, por se preocuparem comigo e cuidarem de mim da melhor forma possível. Vocês são muito maravilhosos!

Agradeço ao meu afilhado Carlos Alberto e meu primo Eduardo por brincarem comigo, me fazerem feliz com as coisas mais simples e mostrarem que o amor faz a vida valer a pena! Amo vocês!

Agradeço também a minha comadre Carla e aos meus queridos padrinhos Cleide e Dodora por estarem sempre presentes em minha vida e por mostrarem que os laços de afeto vão além dos laços de sangue.

Agradeço a Helena minha amiga de uma vida inteira e ao meu grande amigo Alexandre, por serem mais irmãos do que amigos e ainda sim serem grandes amigos.

Meu agradecimento mais que especial a Liliane, por ser o maior presente que recebi nos últimos 2 anos, por me ajudar a segurar toda a barra e pressão, por ser amiga, companheira e por nunca permitir que eu me sentisse sozinha durante o mestrado. Te amo muito Lili!

Agradeço também imensamente ao meu amigo Marcos que me apoiou em todos os momentos, ajudou em todos os desafios e com certeza me tornou alguém melhor. Obrigada por esse companheirismo sem medidas!

Não menos importante, meus agradecimentos ao amigo Jayme e aos amigos do GT-BAVi, João Paulo, João Victor, Jorão, Marcelo, Marcos e Nícolas sem vocês esse trabalho não seria possível. A colaboração, o companheirismo e o suporte de vocês me mostrou o quanto uma equipe de trabalho pode se tornar uma família com o esforço de todos.

Aos grandes amigos que o Lapic me concedeu, Marina e Thomás pessoas maravilhosas e amigos que quero levar para a vida inteira. Obrigada por tudo, vocês são demais!

Aos amigos do mestrado Bruno, Marco Aurélio, Renan e Thiago, obrigada por me aturarem, consolarem e permitirem que o mestrado fosse um lugar melhor com a amizade e trabalho de vocês. Vocês são muito especiais.

Agradeço também aos professores Marcelo Moreno, Alex Borges, Regina Maria e Raul Fonseca, por serem excelentes profissionais e tornarem o mestrado um ambiente de pesquisa e sobretudo um lugar que permite que as pessoas cresçam em conhecimento.

Minha eterna gratidão aos meus professores e orientadores Eduardo Barrére e Jairo Francisco, vocês não só tornaram possível esse trabalho, como também me permitiram viver a experiência do mestrado de maneira comprometida e aprendendo um pouco mais a cada dia. Vocês me ensinaram lições importantes que vão muito além de conhecimento acadêmico. Obrigada pela paciência, dedicação e por acreditarem em mim para o desenvolvimento deste trabalho.

Meu agradecimento aos membros da banca examinadora por terem aceitado o convite.

Agradeço também aos desafios que me tiraram do comodismo e me fizeram mais forte.

Agradeço a Universidade Federal de Juiz de Fora pela oportunidade de crescer em inteligência e sabedoria.

Agradeço a RNP (Rede Nacional de Ensino e Pesquisa) e a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro.

A todos o meu sincero agradecimento!

"O segredo do sucesso é a constância do propósito."

Benjamin Disraeli

RESUMO

Com o acúmulo de informações digitais armazenadas ao longo do tempo, alguns esforços precisam ser aplicados para facilitar a busca e indexação de conteúdos. Recursos como vídeos e áudios, por sua vez, são mais difíceis de serem tratados por mecanismos de busca. A anotação de vídeos é uma forma considerável de resumo do vídeo, busca e classificação. A parcela de vídeos que possui anotações atribuídas pelo próprio autor na maioria das vezes é muito pequena e pouco significativa, e anotar vídeos manualmente é bastante trabalhoso quando trata-se de bases legadas. Por esse motivo, automatizar esse processo tem sido desejado no campo da Recuperação de Informação. Em repositórios de videoaulas, onde a maior parte da informação se concentra na fala do professor, esse processo pode ser realizado através de anotações automáticas de transcritos gerados por sistemas de Reconhecimento Automático de Fala. Contudo, essa técnica produz textos ruidosos, dificultando a tarefa de anotação semântica automática. Entre muitas técnicas de Processamento de Linguagem de Natural utilizadas para anotação, não é trivial a escolha da técnica mais adequada a um determinado cenário, principalmente quando trata-se de anotar textos com ruídos. Essa pesquisa propõe analisar um conjunto de diferentes técnicas utilizadas para anotação automática e verificar o seu impacto em um mesmo cenário, o cenário de similaridade entre vídeos.

Palavras-chave: Recuperação de Informação. Repositório de Videos. Categorização de Texto Ruidoso. Processamento de Linguagem Natural.

ABSTRACT

With the accumulation of digital information stored over time, some efforts need to be applied to facilitate search and indexing of content. Resources such as videos and audios, in turn, are more difficult to handle with by search engines. Video annotation is a considerable form of video summary, search and classification. The share of videos that have annotations attributed by the author most often is very small and not very significant, and annotating videos manually is very laborious when dealing with legacy bases. For this reason, automating this process has been desired in the field of Information Retrieval. In video lecture repositories, where most of the information is focused on the teacher's speech, this process can be performed through automatic annotations of transcripts generated by Automatic Speech Recognition systems. However, this technique produces noisy texts, making the task of automatic semantic annotation difficult. Among many Natural Language Processing techniques used for annotation, it is not trivial to choose the most appropriate technique for a given scenario, especially when writing annotated texts. This research proposes to analyze a set of different techniques used for automatic annotation and verify their impact in the same scenario, the scenario of similarity between videos.

Keywords: Information Retrieval. Video Repository. Noisy Text Categorization. Natural Language Processing.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ASR Automatic Speech Recognition

BBC British Broadcasting Corporation

GT-BAVi Grupo de Trabalho de Busca Avançada por Vídeos

HMM Hidden Markov Models

NER Named Entity Recognition

OCR Optical Character Recognition

PLN Processamento de Linguagem Natural

RI Recuperação de Informação

RNP Rede Nacional de Ensino e Pesquisa

SRL Semantic Role Labeling

TF-IDF Termo Frequency-Inverse Document Frequency

TVSM Topic-based Vector Space Model

WER Word Error Rate

SUMÁRIO

1 INTRODUÇÃO

Atualmente, a Recuperação de Informação (RI) é um tema de grande relevância, principalmente ao considerarmos a quantidade de mídias que são produzidas e compartilhadas a cada instante na internet. Segundo Baeza-Yates and Ribeiro-Neto (2011), os sistemas de RI tem como objetivo recuperar texto, imagem, vídeos e sons relacionados ao conteúdo de interesse do usuário, classificando-os de acordo com a relevância da busca. De todas essas mídias, o vídeo é a que tem o maior destaque, seja pela sua capacidade de adaptação às diversas plataformas quando o assunto é educação (OLIVEIRA et al., 2010) ou mesmo por sua atratividade natural, unindo fatores auditivos e visuais numa única mídia (FIGUEIREDO et al., 2015; MEDEIROS; PANSANATO, 2015).

Estudos na área de RI são importantes, pois a maioria dos usuários ainda não consegue extrair o máximo potencial em suas buscas, sobretudo quando estão relacionadas a vídeos, devido ao fato da linguagem natural ser muitas vezes vaga e incerta, além de existir uma grande dificuldade de representação semântica do seu conteúdo (GUPTA et al., 2015). Em se tratando dos vídeos disponíveis na web, isso se torna ainda mais evidente, pois existem grandes demandas para solucionar esse problema, o que atrai um amplo interesse entre os pesquisadores (JIANG et al., 2013).

Atualmente, diversas técnicas podem ser utilizadas para indexação de vídeos, como classificação por histogramas de cores, formas, reconhecimento de ações, reconhecimento de faces, extração de texto através de Reconhecimento Óptico de Caractere (OCR - *Optical Character Recognition*), entre outras (SOUZA et al., 2017). Independente da forma como os dados são coletados dos vídeos, esses vídeos podem ser associados a conceitos, os quais representam os principais assuntos abordados em seu conteúdo. A este processo é atribuído o nome de anotação semântica (RAMALHO, 2000). Os vídeos anotados semanticamente são relacionados a entidades, que, por sua vez, fazem parte de uma rede de relações entre significados, como uma ontologia ou um tesauro. Sendo assim, a anotação de mídias facilita o processo de busca em repositórios e muitos pesquisadores tem trabalhado em melhorias nesse processo para diversas mídias (DASIOPOULOU et al., 2011; QAZI; GOUDAR, 2016).

Em um cenário mais específico, como o cenário educacional, esse processo se torna

ainda mais importante, pois nem sempre o aluno tem amplo conhecimento sobre o tema e seus termos correlatos. Além disso, muitas das vezes a interface de busca não lhe oferece o apoio necessário para que tal busca tenha resultados satisfatórios. Vale também destacar que o sucesso em encontrar o vídeo relacionado ao tema que está sendo buscado muitas vezes pode proporcionar um aprendizado mais completo para o aluno (NASH, 2005).

Barrére (2014) mostra que um tipo de vídeo que recebe destaque nesse cenário educacional é a videoaula, que se apresenta em grande volume, seja em plataformas educacionais ou mesmo repositórios populares de vídeos. As videoaulas, em sua grande maioria, possuem conteúdo informativo uniforme, dificultando assim a identificação de eventos e a detecção de limite de tomadas (TASKIRAN et al., 2006). Esse fator pode limitar as possibilidades de uso de algumas técnicas de indexação, pois cada parte do vídeo se mostra igualmente importante para o usuário. Nesse tipo de vídeo, Gravier et al. (2015) afirmam que a fala, a linguagem e o áudio têm-se apresentado como importantes portadores de semântica. Em particular, a linguagem é de extrema importância para a compreensão da essência da mensagem. Embora existam variações de videoaulas, é frequente o uso de vídeos reproduzindo aulas num formato mais próximo do tradicional, onde existe um professor (que pode ou não estar visível) discursando sobre um conteúdo e apresentando exemplos ao longo do vídeo.

A maior parte dos buscadores atuais tem como foco a indexação textual (CROFT et al., 2010). Usualmente, esses sistemas de buscas, quando necessitam retornar mídias contínuas, como vídeos por exemplo, utilizam metadados, ou seja, informações textuais que são anexadas ao redor da mídia. O problema relacionado a esses metadados é que embora sejam uma solução aparentemente eficaz para buscas textuais, normalmente são adicionados manualmente, sendo necessário alocar recursos humanos para a classificação, o que torna a abordagem ineficiente, principalmente, para volumosas bases de dados legadas. Outro problema relacionado à anotação manual é que ela depende da experiência, relacionada à narrativa, da pessoa que irá fazer a classificação e isso acaba comprometendo sua eficácia (POLYVYANYY, 2007; TURNBULL et al., 2008).

Uma solução para conseguir outras informações textuais relacionadas a um vídeo, sem depender da intervenção humana, e que vem apresentando resultados relevantes e adapta-se bem a esse tipo de cenário, é a transcrição automática do áudio associado ao vídeo (RAIMOND; LOWIS, 2012; VINCIARELLI, 2005). Porém, essa técnica gera textos

ruidosos (VINCIARELLI, 2005). Entretanto, com o uso de bons modelos de transcrição, essa tarefa já é viável (RAIMOND; LOWIS, 2012). Neste trabalho, denota-se ruído todo o erro gerado pelo processo de transcrição automático de fala. Estes erros geralmente estão associados à existência de homofonias, dificuldade no processamento de áudio de falantes com diferentes sotaques, timbres de voz, etc.

Existem também muitos fatores que tornam a recuperação de texto mais oportuna que a recuperação de imagem, áudio ou de vídeo. O texto traz as palavras como base e suas pontuações são claramente delimitadores na estrutura. Por outro lado, em áudio e vídeo o fluxo de diversas informações é contínuo e os limites estão em sua maioria implícitos. O tamanho dos arquivos é também outro fator em que a diferença é notória, pois arquivos de texto, na maioria das vezes, são bem menores que os demais, influenciando assim nos recursos gastos para manipulação e processamento (BAEZA-YATES; RIBEIRO-NETO, 2011).

Unindo todos esses fatores e considerando o fato que é possível, através do áudio, utilizar as vantagens da recuperação de texto, recorrendo a sistemas de Reconhecimento Automático de Fala (ASR - *Automatic Speech Recognition*) (VINCIARELLI, 2005), o processo se torna ainda mais atraente.

A principal hipótese desse trabalho é que quando anota-se textos ruidosos, as abordagens comuns não são as mais eficientes. Sendo assim, dado um conjunto de abordagens automáticas para anotação semântica envolvendo também combinações de abordagens, o problema central consiste em descobrir quais abordagens possuem melhor eficiência quando o texto é ruidoso. Para realizar essa análise, as abordagens serão aplicadas a um cenário de busca de similaridades entre vídeos.

A relevância dessa pesquisa está no fato que muitos métodos para tratar textos ruidosos são propostos, porém não existe uma análise de qual deles apresenta melhor eficiência. Esta pesquisa então propõe construir essa análise comparativa para um conjunto de abordagens automáticas de anotação semântica, de forma a demonstrar quais geram resultados mais eficientes e qual eficiência dessas abordagens no processo de encontrar similaridade entre vídeos.

Esta dissertação está estruturada da seguinte forma. O Capítulo 2 apresenta o domínio envolvendo a anotação automática de textos com ruídos, explorando as tarefas de processamento de linguagem natural, o processo de anotação, os sistemas de reconhecimento de

fala e os trabalhos relacionados. O Capítulo 3 define o cenário no qual esse trabalho está inserido. O Capítulo 4 apresenta os testes experimentais tanto para a análise de eficiência das abordagens em si, quanto para as abordagens quando aplicadas à finalidade específica de similaridade, e os resultados obtidos. Por fim, o Capítulo 5 apresenta as conclusões do trabalho, contribuições e os trabalhos futuros.

2 ANOTAÇÃO AUTOMÁTICA DE TEXTOS RUIDOSOS

Em repositórios de videoaulas, é comum existir vídeos que representam aulas que simulam uma aula presencial. Geralmente um professor transmite e detalha conhecimento sobre assuntos diversos de forma oral e contínua. Nesse tipo de vídeo, onde todo o conteúdo é igualmente importante do início ao fim, grande parte da informação contida no vídeo é transmitida através do áudio, surgindo a necessidade de processar a informação através do reconhecimento da fala (GRAVIER et al., 2015).

Neste capítulo, serão abordados, na seção 2.1, os principais conceitos de Processamento de Linguagem Natural (PLN), destacando as principais tarefas e dificuldades desta área. Nas seções 2.2 e 2.3, são discutidas as tarefas de anotação e de reconhecimento automático de fala, respectivamente, como tarefas de processamento de linguagem natural. Por fim, os principais trabalhos relacionados com esta pesquisa são apresentados e discutidos na seção 2.4.

2.1 PROCESSAMENTO DE LINGUAGEM NATURAL

Quando considera-se a fala para coletar informações, ingressa-se na área de PLN. O PLN consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural (YE, 2016), como por exemplo: tradução e interpretação de textos, busca de informações em documentos, produção de texto em um formato desejado, manipulação de textos para extração de conhecimento, indexação automática, sumarização, entre outras (CHOWDHURY, 2003; CHOPRA et al., 2013).

Chowdhury (2003) cita que qualquer tarefa de PLN envolve uma questão muito importante, a compreensão da linguagem natural. O autor também afirma que compreender a linguagem natural envolve três grandes problemas: o processo de pensamento, a representação e significado da língua e conhecimento prévio de domínio. Esses problemas tornam a compreensão da linguagem natural uma tarefa difícil (CORRÊA, 2016).

Contudo, a área de PLN busca preparar o computador para lidar com a linguagem

humana. Através da criação de sistemas que possam, mesmo que de maneira simples, simular o conhecimento e o desempenho linguístico dos humanos (CORRÊA, 2016).

Liddy (1998) apresenta os níveis linguísticos e garante que, para entender as línguas naturais, é necessário ser capaz de diferenciá-los. Esses níveis são os mesmos usados pelas pessoas para encontrar o significado de um texto ou idioma. São eles:

- fonético ou fonológico, ou seja, o que trata a pronúncia das palavras.
- morfológico, o qual se preocupa com as partes que compõem as palavras, como a raiz, o sufixo e o prefixo.
- lexical, o qual lida com o significado lexical das palavras e partes da análises da fala.
- sintático, o qual lida com a gramática e estrutura das frases.
- semântico, o qual trata do significado das palavras e frases.
- de discurso, o qual trata da interpretação de um texto formado por vários trechos de palavras e o significado desse texto.
- pragmático, o qual se preocupa em compreender os usos intencionais da linguagem.

Liddy (1998) comenta que esses níveis podem ser usados em sua totalidade ou de acordo com a necessidade. Porém, quanto mais desses níveis forem considerados e aplicados em um sistema, melhor tende a ser o resultado final gerado por esse sistema.

Assim como a linguística possui diversos níveis para a compreensão da linguagem natural, sendo esses níveis interdependentes (CHOWDHURY, 2003), é comum que muitas tarefas de PLN se tornem pré-requisito para a realização de outras tarefas (DURAN, 2016).

Para processar uma língua escrita, é necessário que a máquina entenda o alfabeto, os limites de suas unidades lexicais. Em PLN essa tarefa pode ser vista como tokenização, que permite identificar no texto as unidades estruturais mínimas, como, por exemplo, as palavras (DURAN, 2016). Assim como o *chunker*, um reconhecedor de *chunks* (qualquer conjunto de duas palavras ou mais que possui um significado próprio e só pode ser interpretado em conjunto) que é capaz de identificar no texto palavras compostas (ABNEY, 1991). Essas são tarefas de PLN que se preocupam com o nível lexical da linguística.

Além disso, é importante que a máquina seja capaz de reduzir as diversas formas flexionadas existentes em um documento aos seus radicais. Em PLN essa tarefa é tratada

pela lematização (DURAN, 2016) o processo de normalização onde para cada forma de palavra flexionada em um documento atribui-se a sua forma básica, o lema (KORENIUS et al., 2004) um exemplo é mapear formas de verbo para o tempo infinitivo e as formas dos substantivos para o singular. Essa tarefa também pode ser tratada pela stemmização (FIGUEIREDO, 2017) explorando o nível linguístico morfológico é possível através da tarefa de stemmização, uma análise mais simples do texto, visto que várias palavras são reduzidas se tornando um mesmo radical.

É necessário ainda que o computador seja capaz de rotular essas unidades lexicais de acordo com a sua classificação morfossintática. Em PLN, essa tarefa é conhecida como POS tagging (DURAN, 2016) e identifica na estrutura do texto a função de cada palavra, como os verbos, substantivos, adjetivos, entre outros (SHELKE et al., 2017).

Outra tarefa que indiretamente faz o processo de reconhecer os substantivos é a tarefa de Reconhecimento de Entidade Nomeada (NER - *Named Entity Recognition*). Essa tarefa tem a função de encontrar e classificar nomes de pessoas, lugares e organizações, em textos de linguagem natural. Essa tarefa encontra informações de valor semântico sobre o conteúdo do texto (AMARAL, 2013; SILVEIRA et al., 2015).

A máquina necessita também reconhecer os limites das sentenças (DURAN, 2016) ou ainda detectar *stopwords*, palavras que, isoladas, não definem contexto semântico, como preposições, artigos, entre outros (HOAD; ZOBEL, 2003).

Após realizar essas tarefas mais básicas, é possível que o PLN realize tarefas mais complexas, como atribuir rótulos de análise sintática (tarefa conhecida como *parser*), atribuir rótulos de papéis semânticos (tarefa de anotação de função semântica – SRL - *Semantic Role Labeling*), entre outras (DURAN et al., 2013; DURAN, 2016). A máquina lida assim, com as dificuldades de encontrar identificação para textos, deparando-se com os níveis linguísticos, do mesmo modo que os falantes de uma língua.

2.2 A TAREFA DE ANOTAÇÃO

Na tarefa de anotação semântica automática é preciso distinguir e considerar os níveis semântico e, ao mesmo tempo, lexical e sintático, visto que estão intimamente ligados e dependentes. Um exemplo importante disso são os casos dos sinônimos, palavras diferentes em suas estruturas que possuem o mesmo significado, e das ambiguidades, palavras idênticas em suas estruturas com significados totalmente diferentes (FELDMAN, 1999).

Esses casos são fortes exemplos de situações que podem gerar erros para esse processo, caso ocorram trocas equivocadas durante a tarefa.

A tarefa de anotação semântica, por ser uma tarefa complexa em PLN (DURAN et al., 2013), pode ser executada ou precedida por tarefas mais simples, como as descritas acima. Essas tarefas podem ser realizadas de forma automática por meio de algoritmos que as executem.

É possível que a anotação semântica seja realizada de forma manual, podendo-se utilizar diferentes técnicas para isso. Geralmente a anotação manual é produzida com a ajuda de editores, profissionais que são contratados para indexar informações aos conteúdos que possuem pouco ou nenhum texto associado, como é o caso do repositório de vídeos descrito em (RAIMOND; LOWIS, 2012). Alguns autores propõem processos colaborativos para anotação manual, de forma a atingir um consenso nas anotações (VOLKMER et al., 2005; KAVASIDIS et al., 2014; GRÜNEWALD; MEINEL, 2015). Contudo, a anotação manual é mais custosa e normalmente é feita uma única vez, por ser exaustivo anotar muitas cenas (vídeos) ou parágrafos (texto). Uma forma de diminuir o custo de contratação de mão de obra especializada para anotação é fazer uso de folksonomia (AQUINO, 2007). Ou seja, explorar informações individuais inseridas manualmente em documentos por diversos autores, geralmente de redes sociais ou blogs, para selecionar as melhores anotações para o documento (WU et al., 2006; ELSAYED et al., 2017a,b). Este processo, contudo, depende da existência de documentos pré-anotados e que estes documentos estejam corretamente anotados (XIE et al., 2017). A alternativa menos custosa é a anotação automática, usando das técnicas de processamento de linguagem natural é viável atribuir identificadores de conteúdo nos documentos analisados.

Do ponto de vista da acurácia, anotações manuais tendem a ter melhor resultado, embora anotações automáticas tem alcançado também resultados significativos em diversos cenários, através de diversas técnicas (OLIVEIRA; ROCHA, 2013; SLIMANI, 2013).

2.3 RECONHECIMENTO AUTOMÁTICO DE FALA

Como apresentado na introdução, o processo de anotação semântica automática discutido nesse trabalho está focado no cenário de videoaulas. Neste contexto, a maior parte das informações está presente na fala do professor. Por esse motivo, o processo de anotação semântica automática depende primeiramente de um outro processo, o ASR. Meinedo

(2008) descreve o ASR como um método gerador de transcrições automáticas através do processamento de sinal da fala humana. Segundo o autor, esse método utiliza normalmente 3 modelos em sua composição:

- Acústico: Onde o áudio entra dividido em pequenos pedaços e o modelo acústico do sistema ASR atribui a cada pedaço uma probabilidade ou probabilidade de ocorrência a um fonema determinado.
- Léxico: O modelo léxico determina a sequência de fonemas que formam cada palavra.
- Modelo de linguagem: Esse modelo determina a sequência de palavras de uma determinada língua.

Desse modo, esses modelos precisam considerar principalmente os níveis fonético, morfológico, lexical e sintático. Na prática, a variabilidade acústica e o treinamento dos modelos serão essenciais para o resultado alcançado nas transcrições.

Existem três casos principais de variabilidade acústica que dificultam no momento da transcrição (DAHL, 2015; BENZEGHIBA et al., 2007). O primeiro é a variação do que é dito pelo falante, como mudanças imprevisíveis que ocorrem no falante, o estilo de fala e a pronúncia das palavras. Tudo isso envolve um universo de possibilidades diferentes no que é dito, pois cada pessoa tem uma origem, um costume e uma dicção distinta. Reconhecer essas peculiaridades de todos os falantes possíveis de uma determinada língua se torna impossível e os sistemas de treinamento não são capazes de coletar todas essas informações. O segundo é a diferença que existe nas pronúncias, sotaques e tipos de voz de cada orador, que atinge as mesmas proporções do primeiro caso. E o terceiro são os ruídos que podem ser de circunstâncias e origens diferentes, como, por exemplo, as falhas na gravação do áudio, músicas no ambiente de gravação, falas sobrepostas, falantes simultâneos, entre outros.

Essas variabilidades produzem ruídos como substituições, deleções e inserções de palavras nas transcrições. Além disso, os sistemas ASR são limitados pelas condições de treinamento de seus modelos, como, por exemplo, o modelo léxico, que só pode gerar como saída palavras que pertencem a esse modelo. Se o dicionário não possuir uma determinada palavra, por mais que o falante repita-a, essa palavra nunca será transcrita (VINCIARELLI, 2005).

Todos esses problemas possíveis na etapa de transcrição acabam transmitindo suas consequências para a etapa de anotação que, por sua vez, já possui suas próprias dificuldades, como as citadas anteriormente. Essa união de problemas são erros que precisam ser considerados na anotação semântica automática.

Conceituada a origem dos ruídos e entendidas as limitações dos processos envolvidos, é possível prosseguir para as formas de analisar os textos dos transcritos com a finalidade de encontrar métodos de anotação automática para os vídeos. Na subseção a seguir, serão apresentados os principais trabalhos relacionados a essa pesquisa, o modo como extraem o texto de um vídeo, como lidam com o processamento dos textos extraídos, a geração de anotações automática e as técnicas de PLN utilizadas nesses processos.

2.4 PRINCIPAIS ABORDAGENS PARA ANOTAÇÃO DE VÍDEOS

Os principais trabalhos relacionados a essa pesquisa lidam direta ou indiretamente com o processo de anotação automática. Esses trabalhos utilizam textos gerados através de processo de reconhecimento automático de fala, reconhecimento óptico de caracteres, legendas, entre outros.

A fonte de extração do texto está intimamente ligada com o tipo de ruído produzido. Vinciarelli (2005) relata que o ruído produzido por OCR é diferente do ruído produzido por reconhecimento de voz. Os dois tipos de ruídos impactam de formas diferentes os textos gerados. Os sistemas ASR dão origem a substituições, deleções e inserções de palavras, enquanto os sistemas OCR produzem substituições de palavras ou então se um caractere é reconhecido erroneamente o resultado gerado deixa de ser uma palavra ou expressão válida para o texto.

Sabendo que os ruídos produzidos por fontes diferentes são ruídos diferentes, faz-se necessário o questionamento sobre qual a técnica aplicada na anotação semântica automática traz resultados mais eficientes para lidar com cada tipo de ruído.

Outro fator a se considerar é a finalidade para qual a mídia está sendo anotada, visto que a escolha da técnica impacta no resultado da anotação. É possível, pela escolha da técnica, optar por obter identificadores mais específicos para as partes do documento ou mais amplos para o conteúdo do vídeo como um todo.

Os tipos de técnicas usadas para a tarefa de anotação semântica automática nos trabalhos relacionados foram, em sua maioria, tarefas de PLN citadas anteriormente. Além

dessas tarefas relatadas, estão o modelo de tópicos, o uso de classificadores e a tarefa de *Entity Linking*. O modelo de tópicos foi definido em (BECKER; KUROPKA, 2003) como uma abordagem baseada em vetores para comparação de documentos. Essa técnica pode ou não assumir que os termos são independentes no espaço vetorial. Caso sejam dependentes, aproveita-se das suas relações para não perder a correlação de assunto do documento, onde ganha-se em contexto semântico. Caso não sejam dependentes, ganha-se em nível de processamento, por não considerar essas relações (BECKER; KUROPKA, 2003).

Por sua vez, os algoritmos classificadores realizam o processo de categorização de textos baseados em conjuntos previamente classificados. Essas são técnicas efetivas, porém na maioria das vezes o conjunto de treinamento necessita de um especialista para construí-lo. Por fim, a técnica de *Entity Linking* tem o objetivo de determinar a identidade das entidades mencionadas no texto, para isso, usam bases de conhecimentos externas, de onde são atribuídos os significados para as entidades do documento.

Os trabalhos foram dispostos na Tabela 2.1 e comparados com base nos seguintes parâmetros:

- Abordagem: define o tipo de abordagem utilizada pelos autores para o processo de anotação. É utilizada a classificação de Frank et al. (1999), onde os autores apontam que existem duas maneiras de abordar o problema da anotação. Uma é a atribuição de palavra-chave e a outra é a extração de palavra-chave. A primeira é conhecida também como categorização de texto, onde o documento recebe categorias que o identificam e foi representada na tabela como (1). Na segunda, as palavras do próprio documento são utilizadas para identificá-lo e foi representada na tabela como (2). Esse parâmetro serve para refletir qual tipo de abordagem foi escolhida para o processo de anotação.

- Fonte da informação: define como foi extraída a informação do vídeo para realizar a anotação. A etapa de anotação tem a função de associar identificadores textuais ao vídeo. É necessário entender como foi coletada essa informação textual. As opções seriam: através do áudio, da imagem ou associadas manualmente. Esse parâmetro é necessário para identificar o tipo de ruído produzido durante a extração da informação do vídeo (VINCIARELLI, 2005).

- Ranking: define como os identificadores foram ponderados na etapa de anotação. Algumas técnicas de anotação geram muitos identificadores para o mesmo vídeo. É necessário reconhecer quais desses identificadores representam melhor o conteúdo do vídeo. Para isso são usados técnicas que organizam esses identificadores em ordem de prioridade. Esse parâmetro é importante para entender o que foi levado em consideração na hora de escolher os melhores identificadores para representar o conteúdo.

- Técnicas: define qual técnica foi usada para realizar a anotação. Cada técnica apresenta uma tendência de resultado. Essas técnicas podem ser generalistas e identificarem o assunto geral tratado no vídeo como (RAIMOND; LOWIS, 2012), ou anotarem cada trecho ou palavra de forma independente como (GRÜNEWALD; MEINEL, 2015). Esse parâmetro serve para entender qual técnica é escolhida para cada finalidade.

Nessa tabela 2.1, a coluna denotada com a letra **A** representa a abordagem utilizada pelos autores, a coluna **FI** indica a fonte da informação, a coluna **R** indica a estratégia de ranqueamento e, por fim, a coluna **T** denota a técnica utilizada no trabalho.

Técnicas para anotações semânticas automáticas podem ser usadas para diversos fins, como na indexação de dados da fala a fim de criar uma estrutura de resumo e fornecer ferramentas para navegar nos áudios (MAKHOUL et al., 2000). Em (RAIMOND; LOWIS, 2012), por sua vez, as anotações são usadas para ajudar na navegação e pesquisa dentro dos repositórios de programas da Corporação Britânica de Radiodifusão (BBC - *British Broadcasting Corporation*). Da mesma forma, (KÜÇÜK; YAZICI, 2013; YANG; MEINEL, 2014) também as utilizam para realizar a indexação e recuperação de vídeos. Tu et al. (2014) utilizam as anotações para entender os eventos dos vídeos e responder consultas de usuários. Zhao et al. (2015) constroem um sistema de navegação visual e usam as anotações para criar pistas de navegações visuais e textuais de forma interativa, enquanto Pereira et al. (2017) fazem uso das anotações para o estudo do Português Brasileiro em documentos históricos.

Contudo, é possível levantar duas incertezas, uma está na preocupação de como lidar com a existência de ruído no processo de extração da informação e com o processo de anotação em si. Os trabalhos de Frank et al. (1999); Vinciarelli (2005) se preocupam especificamente com o processo de anotação, não tendo como foco principal a finalidade

TRABALHOS	**A**	**FI**	**R**	**T**
(FRANK et al., 1999)	(2)	Extração de documentos PostScript, OCR	Tf x Idf, a distância de uma frase a partir do início do documento	Stemming, Classificação
(MAKHOUL et al., 2000)	(1)	ASR	hidden Markov models (HMM)	NER, Entity linking (modelo probabilístico)
(VINCIARELLI, 2005)	(1)	ASR, OCR	Tf x Idf	Stemming, Modelo de tópicos
(RAIMOND; LOWIS, 2012)	(1)	ASR	Tf x Idf	Stemming, Modelo de tópicos
(KÜÇÜK; YAZICI, 2013)	(2)	ASR	Frequência das palavras	Stemming, NER, Entity linking (extração de informação)
(YANG; MEINEL, 2014)	(2)	ASR, OCR	Tf x Idf	Stemming, POS tagging (substantivos)
(TU et al., 2014)	(2)	Texto Manual, Detecção de objetos, cenas e eventos	Modelo probabilístico	POS tagging, Chunker, NER, Árvore de parser
(ZHAO et al., 2015)	(1)	ASR, OCR, detecção de cena	TFxIDF	Entity linking
(PEREIRA et al., 2017)	(1)	OCR	Especialista de domínio	Tokenização, Pos Tagging, Entity linking (extração de informação)

Tabela 2.1: Principais trabalhos para anotação de textos com ruídos

para a qual essas anotações serão usadas e em ambos os trabalhos a tarefa de anotação é feita sobre textos com ruído. Vinciarelli (2005) inclusive realiza uma investigação sobre o processo de extração e os possíveis tipos de ruído e seus impactos na anotação.

Dado que o modelo de ASR não gera como resultado um texto perfeito é importante observar que alguns trabalhos se preocupam em realizar tratamento adequado para esse tipo de texto. Esse é o caso dos trabalhos de (VINCIARELLI, 2005; KÜÇÜK; YAZICI, 2013; RAIMOND; LOWIS, 2012). Isso é um fator importante a ser considerado, pois mostra que a preocupação com os ruídos é pertinente.

Vinciarelli (2005); Raimond and Lowis (2012) demonstram isso através da categorização do texto, eles tornam o texto mais genérico através da stemmização e utilizam categorias mais amplas para anotarem o texto. Por sua vez, Küçük and Yazıcı (2013), sabendo das limitações do modelos de ASR para sua linguagem abordada, utiliza um modelo um pouco mais consolidado em outro idioma e realiza a tradução para o seu idioma, usando também a tarefa de stemmização.

Embora tenha-se definido como enfoque do trabalho analisar textos provenientes do ASR, três dos trabalhos não utilizam esse sistema como fonte de extração de texto, são eles (FRANK et al., 1999; TU et al., 2014; PEREIRA et al., 2017). O fato dos trabalhos não usarem o ASR justifica-se por analisarem outros tipos de documentos onde o áudio não é a principal fonte de informação ou é inexistente. Embora esses trabalhos não utilizem o ASR, fazem uso de processos automáticos de extração de informação. Outros tipos de extração automática de texto, apesar de diferirem no tipo de ruído, ainda dão origem a textos ruidosos.

Outra incerteza, é para qual finalidade será utilizada a anotação, pois as tarefas como a abordada em Tu et al. (2014) que utilizam as anotações para entender os eventos dos vídeos e responder consultas de usuários, e a tarefa abordada em Zhao et al. (2015) para construir um sistema de navegação visual e usar as anotações para criar pistas de navegações visuais e textuais de forma interativa ou ainda a tarefa de fornecer ferramentas para navegar nos áudios como em (MAKHOUL et al., 2000), necessitam de anotaçoes muito mais ligadas aos trechos do documento. Portanto, requerem técnicas anotadoras que consigam esmiuçar os documentos para realizar anotações em várias partes do documento. Enquanto, abordagens para realizar a indexação e recuperação de vídeos e áudios, como em (KÜÇÜK; YAZICI, 2013; YANG; MEINEL, 2014; RAIMOND; LOWIS, 2012), necessitam

de abordagens anotadoras que gerem anotações mais gerais sobre os assuntos tratados em todo o documento.

Apesar dos textos terem origens de fontes diferentes e atenderem finalidades distintas, é perceptível que as etapas de pré-processamento se repetem, como a tarefa de stemmização e o POS tagger (que localiza verbos e substantivos). Técnicas muito comuns como o *stopping* e *stemming* são importantes quando é usado um ASR, visto que o resultado desse tipo de processo é um texto contínuo, não capitalizado e sem nenhuma estrutura de pontuação. Quanto ao OCR, por nem sempre possuir um modelo léxico, pode dar origem a qualquer sequência de caracteres, mesmo que esses caracteres não gerem uma palavra (VINCIARELLI, 2005). Esse fato requer atenção, pois o uso de algumas técnicas pode não ser a melhor opção dependendo da fonte de extração do texto.

É possível observar que, apesar da grande variação de tempo entre os trabalhos, não foi consolidado apenas uma forma de abordar a tarefa de anotação, visto as especificidades de cada cenário. Alguns pesquisadores decidem usar as palavras do próprio texto como identificador e outros decidem atribuir outros identificadores que não aparecem especificamente no texto. Quanto à forma de ranquear os melhores identificadores de um documento, muitos trabalhos utilizam o TF-IDF *(termo frequency-inverse document frequency)*, o que indica que a frequência da palavra no texto reflete o fato dessa palavra ser parte da semântica desse texto ou não.

Os trabalhos relacionados foram apresentados em ordem cronológica e é possível observar que existe um período de aproximadamente 18 anos entre o primeiro e o último trabalho disposto na Tabela 2.1. As técnicas utilizadas pelos autores, no entanto, se mantiveram constantes, apesar de existirem pequenas modificações e novas preocupações ao longo do tempo.

É válido observar que, além do processo em si, partes adjacentes do processo de anotação semântica automática sofreram mudanças relevantes, como é o caso das bases de conhecimento, que crescem e se aperfeiçoam ao longo desses anos. Um exemplo disso é a DBpedia, que destaca-se como uma das fontes de conhecimento mais populares e amplamente utilizada para consultas a dados ligados (LEHMANN et al., 2015). A grande vantagem está no fato das bases de conhecimento crescerem à medida que mais informações são inseridas. Dessa forma, os sistemas que as utilizam podem se tornar mais específicos ou mais abrangentes ao longo do tempo. Assim, gera-se resultados positivos

para o processo de anotação.

Entendendo os domínios envolvendo os trabalhos relacionados, observando suas motivações e conhecendo suas soluções, esse trabalho pretende responder as incertezas levantadas nesta seção, através de experimentações em um ambiente controlado com o uso de técnicas de anotações distintas.

3 CENÁRIO DE APLICAÇÃO DOS EXPERIMENTOS

Técnicas para anotações semânticas podem ser utilizadas em diferentes tipos de aplicações com diferentes finalidades. Esse fato faz com que soluções aplicadas a um contexto não sejam adequadas para outros. Não é trivial, contudo, a escolha das técnicas mais adequadas, principalmente quando estas são utilizadas em aplicações que fazem uso de textos que podem conter ruídos. Neste capítulo, é discutido o cenário de aplicação que motivou os experimentos realizados para esta pesquisa, o Grupo de Trabalho de Busca Avançada por Vídeos (GT-BAVi). Além disso, é detalhada a abordagem desenvolvida para identificação de similaridade entre vídeos e sua dependência das anotações automáticas.

3.1 GT-BAVI: BUSCA AVANÇADA POR VÍDEOS

O cenário de aplicação é um cenário real e está contextualizado dentro do Grupo de Trabalho de Busca Avançada por Vídeos (GT-BAVi) da Rede Nacional de Ensino e Pesquisa (RNP)[1]. O objetivo do GT é o desenvolvimento de um protótipo para facilitar enriquecer semanticamente os repositórios de vídeos da RNP e facilitar a busca destes conteúdos. A solução desenvolvida pode ser dividida em 3 passos: a transcrição do áudio dos vídeos, a anotação automática e a recomendação de conteúdo. A Figura 3.1 representa os passos dessa solução. Cada passo é implementado por um módulo correspondente. Embora a solução abranja diversos módulos dentro da arquitetura do protótipo, serão discutidos neste trabalho apenas os três módulos que representam os três passos da Figura 3.1.

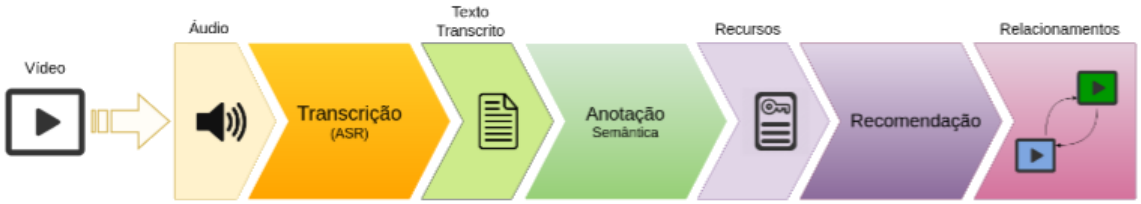

Figura 3.1: Fluxo de trabalho do protótipo do GT-BAVi

[1]https://www.rnp.br

Dentre os repositórios da RNP, o repositório do serviço videoaula@RNP[2] é utilizado como foco para a presente pesquisa, pois possui uma maior uniformidade de temas abordados, fato que maximiza o desempenho da etapa de transcrição. Outro ponto relevante para a escolha desse repositório é fato do mesmo possuir domínios específicos, o que pode facilitar a categorização desse conteúdo.

Foi realizada uma análise dos metadados pré-existentes no repositório da RNP associados aos vídeos. Essa análise foi relevante para uma validação inicial do protótipo. Estes metadados são *tags* associadas manualmente pelos editores dos vídeos para garantir que o vídeo seja encontrado através das buscas por palavras-chave. As informações coletadas para essa análise foram a quantidade de videoaulas dentro do repositório, a quantidade total de metadados utilizados no repositório, quantos desses eram distintos, quantos identificadores em média cada videoaula recebia, a quantidade de vídeos com identificadores inúteis, como identificadores fora do contexto ou muito genéricos. Um exemplo disso são identificadores do tipo: videoaula, videoaulas, *tag*, teste, descrição do professor. As informações coletadas estão dispostas na Tabela 3.1.

Informações	**Valores coletados**
Número de videoaulas	858
Total de identificadores	2225
Total de identificadores distintos	849
Média de identificadores por vídeo	2.59 ± 1.34
Número de videoaulas com identificadores inúteis	604
Número de videoaulas que não possuíam identificadores	2

Tabela 3.1: Cenário da RNP utilizando apenas metadados associados manualmente

O repositório da RNP possuía, no momento da análise, 858 videoaulas. Cada vídeo tinha em média de 2 a 3 metadados, o que totalizava 2225 metadados no repositório. Porém, apenas cerca de um terço desses eram únicos, indicando que muitos dos metadados da base se repetiam. Esse fator torna os resultados das buscas dos usuários no repositório menos exclusivos, pois é possível que muitos vídeos sejam retornados como resultado para uma mesma palavra-chave, quando os termos utilizados para as buscas são mais genéricos. Dos 2225 metadados, 604 apresentam pouca utilidade, ou seja, não acrescentavam identificação específica para os vídeos aos quais estavam anexados. Uma observação importante é que, apesar de 604 vídeos possuírem metadados pouco relevantes, essas videoaulas também poderiam conter identificadores significativos.

[2]http://www.videoaula.rnp.br

Coletando informações mais específicas a respeito dos metadados, encontrou-se também os seguintes dados: 540 videoaulas possuíam apenas 2 metadados, sendo eles "videoaula"e "videoaulas", tornando impossível identificar qualquer um desses vídeos pelo conteúdo. Um dos vídeos possuía apenas o metadado "teste". Ainda, 16 videoaulas possuíam apenas o metadado "Descrição do professor". Esses dados mostram ser impossível encontrar um assunto específico em qualquer um desses vídeos somente pela busca por termos. Os dados coletados apontam também que muitas vezes os metadados anexados ao vídeo de forma manual limita o potencial de uma busca no repositório. Esta situação ocorre principalmente pelo informalismo e pouca dedicação durante a etapa de geração dos metadados das videoaulas, gerando metadados inadequados para uma futura busca pela videoaula.

Os dados da análise mostraram que, em alguns cenários, como o repositório de videoaula da RNP, os metadados que são associados manualmente às mídias podem não agregar o valor desejado, seja por descuido do editor do conteúdo ou por desconhecimento da importância desses metadados. Para permitir melhores resultados, as anotações automáticas passam a ser úteis, mesmo existindo a possibilidade de erro agregado ao processo automático. Esse erro, por sua vez, quando controlado, pode não ter uma influência tão negativa na busca, uma vez que os metadados atribuídos manualmente também possuem erros ou muitas vezes não fazem sentido.

No intuito de melhorar os metadados, associando novos identificadores aos vídeos para aumentar as chances de sucesso nas buscas nos repositórios de videoaulas, foram escolhidas algumas abordagens de anotação semântica automática. Esse conjunto de abordagens foi escolhido por serem opções usuais na literatura, como aponta o estudo apresentado na seção 2.4. Além disso, as abordagens escolhidas possuem níveis de complexidade distintas, desde opções mais simples como *Chunker*, até opções mais sofisticadas como Modelo de Tópicos (TVSM - *Topic-based Vector Space Model*), passando por opções muito usadas como o NER e o *Entity linking*. Desta forma, é possível avaliar o custo-benefício de cada opção.

Como a maior parte dos vídeos dentro do repositório possuem metadados muito fracos e seus títulos também não representam o assunto tratado, como, por exemplo, títulos muito genéricos como: "Exercicio_5". A opção escolhida pelo GT-BAVi para extrair outros textos do vídeo foi a transcrição automática com o uso de um ASR. Essa escolha justifica-se por

se tratar de videoaulas, onde todo o assunto do vídeo se mostra igualmente importante e concentrado na fala do professor.

Um modelo de ASR ideal reconheceria todas as palavras de um determinado idioma, ditas por qualquer pessoa, em qualquer condição de ruído ambiente. Contudo, os sistemas hoje para o português do Brasil ainda apresentam desempenho insatisfatório, quando lidam com sotaques diferentes e com diferentes tipos de ruído e distorção no sinal de fala (OLIVEIRA et al., 2012). Uma alternativa com chance de resolver o problema de baixo desempenho de um ASR é treiná-lo para as condições em que irá operar. Para isso, as bases de dados devem contemplar o cenário alvo da aplicação final. No GT-BAVi é realizado o reconhecimento do que é falado em videoaulas gravadas por falantes do português brasileiro. Portanto, é natural treinar o sistema com amostras de videoaulas gravadas nas mesmas condições em que se pretende rodar a solução proposta pelo GT.

Para a avaliação do sistema de reconhecimento de fala desenvolvido no GT, usou-se a taxa de erro de palavras (WER - *Word Error Rate*), a qual representa a quantidade de modificações que são necessárias em uma sentença reconhecida para transformá-la em uma sentença real. Essas modificações podem ser alterações como inserção, substituição ou remoção de palavras (OLIVEIRA et al., 2012). Os resultados da WER obtidos pelo processo de transcrição criado pelo GT estão próximos dos obtidos em sistemas comerciais como Google[3], Microsoft[4] e IBM[5]. Para a base de videoaulas obteve-se 44,8% de WER enquanto o do Google obteve 35,9%, o da IBM 73,7% e o modelo da Microsoft obteve um resultado de 44,7%.

Ainda que a taxa de erro de palavras do módulo de ASR seja próxima de sistemas reconhecidos pelo mercado, o módulo de anotação terá que lidar com esses erros. Esse fato torna a escolha do processo de anotação uma tarefa singular e complexa. Principalmente, quando existem diversas técnicas que realizam essa tarefa presentes na literatura. Para o processo de anotação foram estudadas e comparadas algumas técnicas. São essas: NER, *Chunker*, *Entity Linking* e Modelo de tópicos (TVSM). Essas técnicas trabalham de formas distintas e geram resultados distintos como *tags* de um texto. Para exemplificar essas diferenças considera-se o texto abaixo, oriundo de um sistema de reconhecimento automático de fala.

[3]https://cloud.google.com/speech/

[4]https://docs.microsoft.com/pt-br/azure/cognitive-services/Speech/API-Reference-REST/BingVoiceRecognition

[5]https://www.ibm.com/watson/developercloud/speech-to-text.html

```
to keep up that trust and of rising of relatively young top chess
   player challenge and successfully dethroned the reigning world
   champion mikhail plop in a much the same way as aman armenia
   american
```

No exemplo acima, a frase originalmente falada foi *"that Tigran Petrosian arrising a relativelly young top chess player challenged and successfully dethroned the reigning world champion Mikhail Botvinnik. Much in the same way as on Armenian-American"*. Percebe-se os erros existentes, como a troca de "that Tigran Petrosian" por "to keep up that trust"e acréscimos de palavras como "aman". Como forma de exemplificar as técnicas utilizadas neste trabalho, este mesmo trecho de transcrição automática será utilizada nas próximas seções.

3.1.1 CHUNKER

A primeira abordagem a ser considerada para essa análise é o *chunker*. Dentre as abordagens de PLN utilizadas para essa pesquisa essa é a mais simples em nível de complexidade. Sua tarefa consiste em segmentar uma sequência de caracteres de entrada em *tokens*. Esses *tokens* são geralmente palavras, pontuação, números, etc. Os *tokens* são detectados com base em um modelo de probabilidade.

Com base no próprio *token* e no contexto que esse está envolvido é marcado o seu tipo de palavra correspondente como pronomes, substantivos, preposições, entre outros. Para realizar essa tarefa usa-se um dicionário de marcações, com base na gramática do idioma que está sendo analisado. Essa etapa recebe o nome de *POS tagger*.

Após receber as marcações estruturais, o texto é dividido em partes sintaticamente correlacionadas de palavras, como grupos de nomes, grupos verbais, mas não especifica se sua estrutura interna faz sentido, nem qual o seu papel na sentença principal.

O *chunker* buscará no texto combinações de palavras que tenham sentido sintaticamente, como palavras compostas por exemplo, ou mesmo palavras que sozinhas já tenham sentido completo, caso nenhuma outra palavra próxima faça parte do mesmo grupo sintático.

Encontrando os grupos sintáticos de palavras e então os blocos mais relevantes o *chunker* reconhece entidades importantes para o texto. Mesmo sendo uma ferramenta que utiliza a sintaxe para reconhecer *tags* para um texto é importante salientar que muitas

vezes um grupo de palavras como por exemplo “Rio de Janeiro”tem maior valor semântico, que simplesmente “de”.

O *chunker* pode ainda encontrar combinações de palavras que não existam nas bases de conhecimento e então ele as identificará, porém não existirá um recurso correspondente. Nesse caso, nada será anotado em relação ao trecho encontrado.

Para o texto de entrada que foi dado como exemplo, a saída do *Chunker* seria a apresentada abaixo:

```
"player challenge " : " "
"armenia"           : "http://dbpedia.org/resource/Armenia"
"chess"             : "http://dbpedia.org/resource/Chess"
"world champion"    : "http://dbpedia.org/resource/World_champion"
```

O exemplo acima mostra as anotações que o *chunker* encontrou para o texto que foi dado como entrada. Um dos *tokens* encontrados pelo algoritmo não possui correspondente na base de conhecimento que é o caso do *“player challenge”* e por esse motivo aparece no exemplo com o recurso vazio. Entretanto, o algoritmo reconheceu *“player”* e *“challenge”* como substantivos e o fato de ambas as palavras estarem posicionadas uma ao lado da outra possibilitou a interpretação de palavra composta e por esse motivo foram identificadas como um único *token*. Porém, como um ponto positivo, o *chunker* reconheceu *“armenia”* e *“chess”*, *tags* que foram anotadas pelos editores manualmente, o que demonstra que a abordagem é capaz de reconhecer *tags* relevantes para o texto. Quanto ao *token “world champion”* foi encontrado seu correspondente na base de conhecimento, porém essa *tag* não é considerada fundamental para o texto, visto que não representa o assunto principal tratado no programa, esse *token* foi apenas mencionado durante o áudio, mas reconhecido pelo algoritmo.

3.1.2 NER

Essa abordagem compartilha de um processo parecido ao executado pelo *chunker*. A etapa de tokenização e *POS tagger* também ocorre para o NER. Porém, ao invés de reconhecer apenas os blocos de palavras como grupos de nomes, grupos verbais, entre outros, tal qual o *chunker*, posteriormente este processo utiliza também um modelo dependente do tipo de entidade para o qual foi treinado e do idioma. Esse modelo possibilita que o NER encontre no texto grupos de palavras que sejam especificadas em seu modelo como sendo pessoas,

lugares e organizações, para serem anotadas e então as identificará juntamente com seu tipo. Dado o exemplo como texto de entrada, o algoritmo encontrará os resultados a seguir:

```
"mikhail plop"  : " "
"armenia"       : "http://dbpedia.org/resource/Armenia"
"chess"         : "http://dbpedia.org/resource/Chess"
```

O resultado encontrado pelo NER apresenta também um caso semelhante ao da abordagem anterior onde o *token* é reconhecido pelo algoritmo, porém não existe um correspondente na base de conhecimento que é o caso de *"mikhail plop"*. O NER reconheceu o *token "mikhail plop"* como nome de pessoa e por isso identificou como entidade válida. Assim como o chunker a abordagem reconhece recursos que são considerados relevantes para o texto como *"armenia"* e *"chess"*. Apenas *"armenia"* e *"chess"* serão anotados, visto que *"mikhail plop"* não possui um recurso, o que aponta uma característica particular do NER que não se restringe a esse exemplo mas sim a análise de um modo geral, que é a tendência dessa abordagem de anotar poucos recursos para um texto com boa precisão.

3.1.3 ENTITY LINKING

A abordagem *Entity Linking*, assim como as anteriores, realiza o processo de tokenização. Porém, neste caso, também pode remover as *stopwords* e realizar a stemmização do texto como etapas de pré-processamento. Após essa etapa é feita uma ligação desses *tokens* com recursos prováveis de defini-los em uma base de conhecimento. A abordagem então realiza a desambiguação desses recursos candidatos e anota o recurso considerado correspondente.

O modelo de *Entity Linking* anotará menções de recursos de uma base de conhecimento que estiverem presentes no texto. A abordagem de *Entity Linking* usada no GT-BAVi e também nesse trabalho é a descrita em (MENDES et al., 2011). Dado o texto inicial, como exemplo de texto a ser anotado por esse algoritmo, a saída será a demonstrada a seguir:

```
"trust"             : "http://dbpedia.org/resource/Trust_law"
"chess"             : "http://dbpedia.org/resource/Chess"
"aman"              : "http://dbpedia.org/resource/Aman"
"armenia"           : "http://dbpedia.org/resource/Armenia"
"american"          : "http://dbpedia.org/resource/United_States"
```

Ao contrário da abordagem NER, que tende a anotar poucos recursos para um texto, o *Entity Linking* tende a anotar muitos recursos, visto que todos os *tokens* anotados, dependendo da técnica utilizada, provavelmente possuirá um recurso que compartilhará parte do seu nome. Apesar da abordagem reconhecer os mesmos recursos relevantes encontrados pelas abordagens anteriores, essa abordagem reconhece também muitos outros recursos que não são tão relevantes para o assunto tratado, como é o caso de *"trust"*, *"aman"* e *"american"*.

3.1.4 MODELO DE TÓPICOS

O modelo de tópicos, assim como o *Entity Linking* e as demais abordagens, também pode realizar uma etapa de pré-processamento do texto. Contudo, sua função é, dadas as palavras em sua ordem textual, realizar uma representação dessas palavras em tópicos específicos. Assim, o modelo possibilita associar "casa branca"com um significado especial no tópico 'política', mas não no tópico 'imobiliário'.

Por esse motivo, essa técnica pode encontrar recursos que tenham relevância para um texto, mesmo que não tenha sido citado nele. Como a análise feita por esse algoritmo é uma análise global do texto, esse não identifica palavras do texto que são relevantes, mas, sim, recursos relevantes para o texto como um todo. A abordagem de modelo de tópicos utilizada nesse trabalho é a mesma descrita em (RAIMOND; LOWIS, 2012).

Dado o exemplo como texto de entrada, o algoritmo encontrará os resultados a seguir:

```
"http://dbpedia.org/resource/Chess"
"http://dbpedia.org/resource/Armenia"
"http://dbpedia.org/resource/Competition"
```

Ao contrário das abordagens anteriores, que identificam os *tokens* relevantes para o texto e procuram recursos que são correspondentes, o modelo de tópicos analisa o texto como um todo e então encontra tópicos capazes de defini-lo. Um exemplo disso é que ele anotou as mesmas tags relevantes que as outras abordagens anotaram, mas foi além, ao encontrar uma tag que identificava não só algo mencionado pelo falante, como também um dos temas gerais do texto que era competição ("competition").

3.1.5 CONSIDERAÇÕES SOBRE AS ABORDAGENS

Em todas as abordagens foram encontradas as *tags "armenia"* e *"chess"*. Além dessas duas *tags*, seria importante que fosse encontrado também o recurso *"Chess_Olympiad"* que não foi reconhecido por nenhuma das abordagens. Porém, não desvaloriza a anotação como um todo, visto que duas de três *tags* foram encontradas.

Por se tratarem de processos automáticos, as abordagens apresentadas encontram alguns recursos não tão satisfatórios e as vezes não reconhecem todos os recursos considerados relevantes. Contudo, todas as abordagens no exemplo apresentado são capazes de identificar tags válidas, o que mostra a relevância de todas para o cenário. Esse fato confirma que tags anotadas por editores são possíveis também de serem anotadas por algoritmos, entretanto, com menos mão de obra.

Em relação ao fluxo de trabalho do GT-BAVi, as abordagens de anotações descritas neste capítulo são posteriormente processadas pelo módulo de recomendação. Este módulo faz a recomendação através do cálculo de similaridade entre os vídeos. O módulo de recomendação busca por videoaulas que falam do mesmo assunto dentro do repositório e as recomendam para usuários dos repositórios da RNP. A descrição da abordagem de recomendação será descrita na próxima seção deste trabalho.

3.2 RECOMENDAÇÃO DE VÍDEOS ATRAVÉS DO CÁLCULO DE SIMILARIDADE DE DOMÍNIO

Uma vez que um repositório de vídeos seja enriquecido com anotações semânticas, é possível buscar por esse conteúdo fazendo uso das relações entre as entidades anotadas. No contexto do GT-BAVi, espera-se que vídeos de interesse de um usuário possam ser recomendados a ele durante a reprodução de um certo conteúdo. Neste cenário, contudo, não há informações de sessão do usuário que possa guiar algoritmos de recomendação através de histórico de buscas e perfil do usuário. Assim, foi projetada uma solução de recomendação de vídeos através da análise de domínio de cada vídeo. A análise de domínio é feita através de um cálculo de similaridade de domínio que faz uso das anotações associadas a cada vídeo e de informações de domínio que são extraídas da DBpedia[6], projeto que extrai conhecimento estruturado da Wikipedia. O processo como um todo

[6]http://www.dbpedia.org

está descrito em (DIAS et al., 2017).

3.2.1 FORMALIZAÇÃO

Considerando apenas as propriedades que serão usadas no cenário de aplicação, pode-se definir um vídeo v_i como uma tupla $v_i =< r_1, r_2, r_3, ..., r_n >$ onde r_j são recursos da DBpedia, ou seja, URIs que identificam pessoas, lugares e organizações dentro da DBpedia. Os recursos r_j podem ser considerados metadados do vídeo que foram associados através das abordagens de anotação semântica citadas na seção 2.2.

Neste trabalho, um recurso é um identificador da DBpedia, a qual corresponde a uma página da Wikipedia que descreve um assunto qualquer como, por exemplo, "calor"ou "algoritmos". Esse recurso da DBpedia pode ser formalizado como uma tupla $r_j =< uri_j, L(r_j), C(r_j), P(r_j) >$, onde uri_j representa o identificador do recurso, $L(r_j)$ o conjunto dos nomes (label) que esse recurso possui, $C(r_j)$ o conjunto de categorias desse recurso e $P(r_j)$ o conjunto de propriedades do recurso.

A partir de então, considera-se as seguintes definições para este processo, R é um conjunto de todos os recursos associados a um vídeo onde r_j é um recurso específico, na DBpedia inglês, ex.: <http://dbpedia.org/resource/Light>. Como o conjunto de todas as categorias associadas a um desses recursos específico tem-se $C(r_j)$. Por exemplo, para o recurso Light, tem-se:
C(http://dbpedia.org/resource/Light) = {http://dbpedia.org/resource/Category:Light}.

Expande-se então o conjunto de categorias, pegando as categorias $C^+(r_j)$, que é o conjunto de todas as categorias associadas como mais abrangentes que alguma categoria $c \in C(r_j)$. Assim, tem-se que $C^+(r_j) = \{x | x \in up(c, \alpha) \wedge c \in C(r_j)\}$, onde $up(c, i)$ representa as categorias mais abrangentes que c até uma distância α de c no grafo. Para isso consulta-se a propriedade skos:broader. Por exemplo, para o recurso Light, tem-se: C^+(http://dbpedia.org/resource/Light) = {Electrodynamics, Electromagnetic_radiation, Electromagnetic_spectrum, Optics, Wave}.

De forma semelhante, expande-se ainda mais o conjunto de categorias, pegando também as categorias $C^-(r_j)$, que é o conjunto de todas as categorias associadas como uma categoria menos abrangente que uma categoria $c \in C(r_j)$. Formalmente, $C^+(r_j) = \{x | x \in down(c, \beta) \wedge c \in C(r_j)\}$, onde $down(c, i)$ representa todas as categorias menos abrangentes que c até uma distância β de c no grafo. Então consulta-se a propri-

edade skos:broader_of, como, por exemplo, C^{-}(http://dbpedia.org/resource/Light) = {Darkness, Light_sources, Lighting, Photons, Vision, Photochemistry, Light_therapy, Fictional_characters_who_can_manipulate_light}.

Portanto, obtêm-se o conjunto de todas as categorias encontradas para r_j, como sendo $\varphi(r_j) = C^{+}(r_j) \bigcup C(r_j) \bigcup C^{-}(r_j)$. Assim, a lista de categorias relacionadas a um vídeo é dado por $\Upsilon(v) = \bigcup_{r \in v} \varphi(r)$. Então, concluímos que os vídeos v_i e v_j serão relacionados se $sim(v_i, v_j) > \alpha$, sendo $\alpha \in [0, 1]$ um valor pré determinado. A similaridade entre dois vídeos pode ser calculada como uma função genérica sim. As funções de similaridade usadas nesse trabalho são apresentadas na seção 4.3.2.1 por meio de dois métodos de ranqueamento distintos.

A predição de relacionamentos se dá através do cômputo da similaridade de tal forma que os vídeos com maior valor de similaridade serão acrescidos deste relacionamento.

3.2.2 EXEMPLO E ALGORITMO

Para exemplificar a abordagem acima, considere a Figura 3.2. A figura apresenta, de forma mais simplificada, as relações entre vídeos, recursos e categorias. Neste exemplo, tem-se que $v_1 =< r_1, r_4 >$, $v_2 =< r_2, r_3 >$ e $v_3 =< r_1, r_3, r_5 >$. Verificando em relação a v_3, este compartilha exatamente um recurso com v_1 e um recurso com v_2. Contudo, ao analisar as categorias associadas aos recursos, verifica-se que o vídeo v_3 possui mais categorias em comum a v_1 que a v_2.

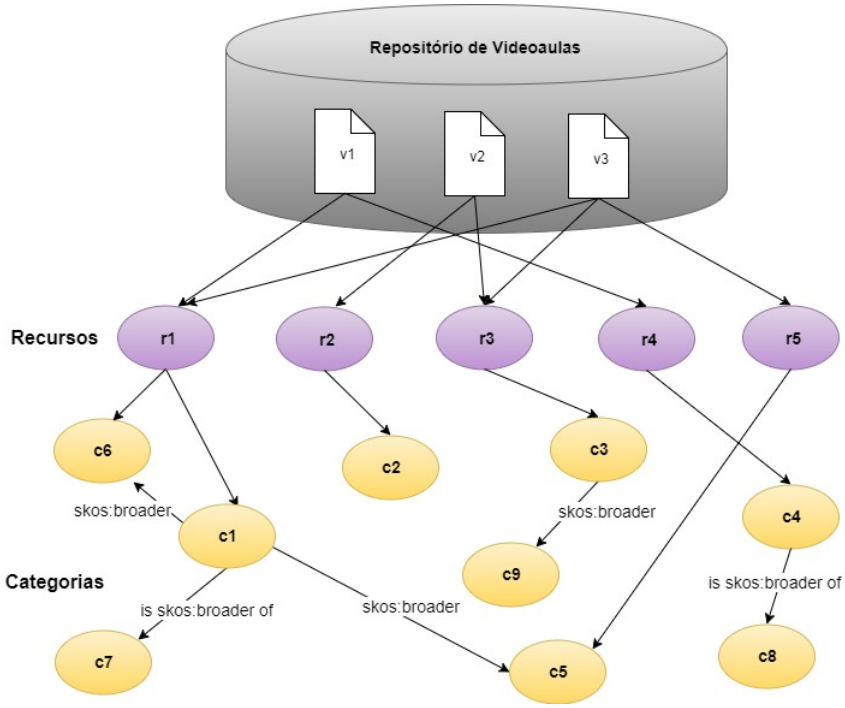

Figura 3.2: Ligações entre conceitos e categorias na DBpedia

Para melhor compreensão da proposta, o algoritmo abaixo representa todo o processo:

Algoritmo 1: Algoritmo para cálculo de similaridade

Entrada: vídeo anotado $v_i =< r_1, r_2, r_3, ..., r_n >$, threshold λ, profundidades α e β

Saída: conjunto R de vídeos relacionados v_i

início
  $Lista \leftarrow \emptyset$
  para *cada* $r \in v_i$ **faça**
    $r \leftarrow getSameAsDBpediaEN\,(r)$
    $\varphi\,(r) \leftarrow$ categorias de r
    $\Upsilon \leftarrow \Upsilon \bigcup \varphi\,(r)$
  fim para
  para *cada categoria* $c \in \Upsilon$ **faça**
    para *cada vídeo* v_j *que possui categoria* c **faça**
      se $sim\,(v_i, v_j) > \lambda$ **então**
        $Lista \leftarrow Lista \bigcup v_j$
      fim se
    fim para
  fim para
  retorna *Lista*
fim

3.2.3 TRABALHOS RELACIONADOS À RECOMENDAÇÃO DE VÍDEOS E AO CÁLCULO DE SIMILARIDADE DE DOMÍNIO ATRAVÉS DO CAMINHAMENTO EM SISTEMAS DE CATEGORIAS

Dados ligados podem ser usados para agregar valor ao conjunto de dados original (ARAÚJO; SOUZA, 2011), desenvolvendo-o com associação a outros conjuntos de dados conexos, além de ser possível utilizá-los de diferentes formas e para diferentes propósitos, como é o caso do Sindice (OREN et al., 2008). Além disso, essas bases externas de conhecimento tem sido utilizadas para apoiar tarefas, tais como anotação semântica (MENDES et al., 2011), alinhamento de entidades (JAIN et al., 2010), identificação de contexto (KAWASE et al., 2014), dentre outras. Neste trabalho é considerado o uso do sistema de categorias da DBpedia para determinar a similaridade entre recursos educacionais anotados semanticamente.

Conforme discutido em (HERRERA et al., 2016), à partir de uma entidade da DBpedia, é possível percorrer diversos caminhos distintos no grafo de relacionamentos para alcançar uma dada entidade. O grau de similaridade entre entidades pode ser medido de acordo com a quantidade de percursos entre duas entidades, a distância desses percursos e os tipos de relações existentes nestes percursos, como é o caso do DBpedia Profiler (HERRERA et al., 2016) e dos trabalhos de Zhu and Iglesias (2017) e de Cheniki et al. (2016).Nestes três trabalhos, diferentes abordagens para encontrar caminhos entre entidades e determinar suas similaridades são propostas. Conforme discutido pelos autores, para um melhor desempenho o cálculo de similaridade necessita processar previamente toda a base da DBpedia, o que exige um reprocessamento a cada atualização da base de conhecimento. Por outro lado, ao fazer uso de todo o grafo da DBpedia, o valor de similaridade entre entidades possui boa acurácia. Assim, é possível analisar a similaridade entre recursos educacionais através da similaridade entre cada par de anotações.

Se os recursos educacionais tiverem outro tipo de informação associada, como texto, abordagens similares são encontradas na literatura. Por exemplo, o sistema de categorias pode ser utilizado para encontrar o tópico abordado pelo recurso educacional, como no trabalho de Hulpus et al. (2013) e Köhncke and Balke (2010). Ambos os trabalhos utilizam informação textual e as relaciona com entidades da DBpedia para, em seguida, processar o grafo de relações entre entidades e identificar o tópico que melhor representa

um documento. Neste processo, ambos os trabalhos precisam tratar de problemas como identificação de recursos candidatos e desambiguação de recursos, visto que um mesmo termo no texto pode representar diferentes entidades.

No cenário tratado nesta dissertação, um vídeo ou videoaula pode ter várias anotações, as quais podem fazer parte de diversas categorias distintas. A similaridade é calculada através das categorias em comum nas quais essas anotações se ligam no grafo da base de conhecimento. Assim, quanto mais os recursos educacionais compartilharem de um mesmo assunto, mais similares eles serão. O uso desse subgrafo é suficiente para gerar bons resultados, o que acarreta uma quantidade menor de dados a serem processados e por consequência, um melhor tempo de processamento.

Com a dificuldade de recuperar informação presentes em vídeos, alguns autores têm se dedicado a usar bases de conhecimento para auxiliar na identificação de similaridade entre videoaulas e outros tipos de vídeos. Neste caso, é comum o uso de texto associado ao vídeo, como títulos, resumos e outros metadados, como também legendas ou tags criadas por usuários. Por exemplo, o trabalho de Chen et al. (2010) que utiliza as categorias da Wikipedia para melhorar a categorização de vídeos identificando conceitos dos vídeos na Wikipedia através das tags e títulos criados por usuários. Dessa forma, os autores exploram as categorias associadas aos conceitos para fazer uma classificação desses vídeos. Contudo, os conceitos da Wikipedia são identificados manualmente a partir de títulos e tags de vídeos. Por outro lado, o trabalho de Raimond and Lowis (2012) utilizam transcrições automáticas do áudio de vídeos para identificar entidades na DBpedia e, com isso, associá-las aos vídeos. Os autores utilizam um modelo de espaço vetorial que possui a função de comparar a similaridade semântica entre expressões. Assim, as categorias da DBpedia são usadas para gerar interpretações para os termos presentes em um determinado transcrito a fim de identificar o contexto correto para as interpretações encontradas. Uma vez que as transcrições automáticas podem gerar muito ruído e as tags podem ser muito vagas, ambas as abordagens possuem problemas com acurácia.

Por fim, o cálculo de similaridade usando bases de conhecimento como apoio auxiliam em diversos tipos de sistemas. Se considerarmos especificamente a recuperação de vídeos educativos, o YoVisto (SACK; WAITELONIS, 2010) é o principal mecanismo de busca neste contexto. Este é um buscador que também utiliza ligações a dados externos para enriquecer as buscas. Sua especialidade está em vídeos acadêmicos e conferências. A

principal contribuição que o YoVisto traz é o fato de usar a indexação do conteúdo com baixa granularidade, segmentando e estabelecendo tags por quadro ou trechos do vídeo. Para extrair informações esse buscador utiliza Folksonomia e técnicas de processamento de imagens, mais precisamente OCR (WAITELONIS et al., 2010). Estas informações são relacionadas com entidades da DBpedia para possibilitar ao motor de busca sugerir conceitos. Já o Iris AI [7], assim como o YoVisto, está voltado para o cenário acadêmico, onde sua função é a recomendação de artigos científicos a partir de uma indicação inicial do usuário. O Iris AI se aproxima desta dissertação quando utiliza a similaridade entre documentos com uso de bases de conhecimento para relacioná-los. O sistema, por sua vez, usa um modelo de recomendação e processamento de linguagem natural a partir dos vídeos do repositório TED Talks [8] para treinar esse modelo. Técnicas de inteligência artificial são parte fundamental do processo e necessita de bastante treinamento para conseguir alcançar boa acurácia.

Após apresentadas as abordagens utilizadas para a anotação semântica automática, demonstrado o algoritmo de busca por similaridade e os trabalhos relacionados ao cenário de aplicação presentes na literatura, será apresentado no próximo capítulo as experimentações e os resultados encontrados.

[7] http://iris.ai
[8] http://www.ted.com

4 EXPERIMENTAÇÃO

Nesta pesquisa dois tipos de experimentos foram realizados. O primeiro tipo de experimento foi realizado em relação a anotação semântica automática. O segundo tipo de experimento realizado foi em relação a similaridade entre vídeos, baseado nas anotações apresentadas no primeiro tipo de experimento. Por fim, os dois tipos são relacionados, com o intuito de observar o impacto de cada abordagem de anotação na busca de similaridade entre vídeos.

Neste capítulo, serão abordados, na seção 4.1, as métricas utilizadas para avaliar a eficiência das abordagens nos dois tipos de experimentos. Nas seções 4.2 e 4.3, são discutidos os experimentos com as abordagens de anotações automáticas e os experimentos com o cálculo de similaridade entre vídeos.

4.1 MÉTRICAS

Nesta seção serão apresentadas as métricas de avaliação que foram utilizadas nos dois tipos de experimentos dessa pesquisa, são elas: a precisão, a cobertura e o topN. Optou-se por essas métricas por serem conhecidas e usadas na literatura como em (MAKHOUL et al., 2000; VINCIARELLI, 2005; RAIMOND; LOWIS, 2012; KÜÇÜK; YAZICI, 2013).

Nas fórmulas de precisão e cobertura interpreta-se como itens_relevantes os identificadores relevantes (no caso da anotação) e os vídeos relevantes (no caso da similaridade). A mesma interpretação pode ser adotada para os itens_recuperados. A precisão foi calculada pela seguinte fórmula:

$$\text{precisão} = \frac{\sum(itens\_relevantes \bigcap itens\_recuperados)}{\sum(itens\_recuperados)}$$

A precisão será máxima (igual a 1) quando todos os itens recuperados forem itens relevantes e mínima (igual a 0) quando nenhum item relevante for recuperado. A métrica de cobertura é calculada pela fórmula a seguir:

$$\text{cobertura} = \frac{\sum(itens\_relevantes \bigcap itens\_recuperados)}{\sum(itens\_relevantes)}$$

A cobertura será máxima (igual a 1) quando todos os itens relevantes forem recuperados, cobrindo todo o espaço dos resultados esperados, e mínima (igual a 0) quando nenhum item relevante for recuperado.

Para calcular o topN os itens retornados foram ranqueados considerando o número de categorias em comum que possuem com o vídeo inicial. Quanto maior for esse número, mais correlatos são os assuntos tratados nos vídeos, justificando uma melhor colocação no ranking. O topN foi calculado através da fórmula adaptada de (BERENZWEIG et al., 2004), a seguir:

$$\text{TopN} = \frac{\sum_{j=1}^{N} \omega^{k_j}}{\sum_{i=1}^{N} \omega^{i}}$$

Na fórmula acima, N é o número de identificadores (no caso da anotação) ou vídeos (no caso da similaridade) manualmente relacionados pelos especialistas, k_j é a posição do identificador j no caso anotação e do vídeo j no caso da similaridade no resultado dos vídeos automaticamente relacionados e $0 < \omega < 1$ é uma constante de decaimento, a qual expressa o quanto se quer penalizar um identificador ou um vídeo por aparecer em posições inferiores no resultado. Para o estudo foi escolhido $\omega = 0.8$, em ambas as avaliações, o que significa que um identificador ou um vídeo irá contribuir com aproximadamente 0.1 no resultado final antes da normalização se ele ocorrer na décima colocação do resultado ranqueado.

4.2 ANOTAÇÃO AUTOMÁTICA

Para a experimentação utilizou-se quatro abordagens distintas de PLN na tarefa de anotação. Realizou-se também combinações entre as abordagens, a fim de verificar se uma agregaria valor a outra. Além disso, foi considerada uma abordagem de anotação simples como uma linha de base (*baseline*) para os outros experimentos.

4.2.1 DESCRIÇÃO DOS EXPERIMENTOS

Para avaliar a tarefa de anotação utilizou-se um *benchmark*[1] de 132 textos transcritos dos programas de rádio da BBC[2], no idioma inglês. Estes transcritos são oriundos dos áudios dos programas de rádio da BBC produzidos por um ASR com modelos bem treinados. Em consequência desse bom treinamento, obtêm-se textos com boa qualidade de transcrição, próximo da taxa de erro de palavra encontrada pelo GT-BAVi. Os programas transcritos são programas de notícias, onde o conteúdo, assim como nas videoaulas da RNP, são

[1] https://github.com/bbc/automated-audio-tagging-evaluation
[2] http://bbc.co.uk/programmes

conteúdos informativos uniformes. Além dos transcritos, o repositório possui anotações manuais realizadas por editores contratados pela BBC. Assim, algoritmos de anotação automática podem ser avaliados através da comparação com as anotações manuais.

Optou-se por esses transcritos por serem textos gerados com boa qualidade, apesar de possuírem ruídos como todos os textos originados de ASR. Esses textos foram anotados por profissionais, o que possibilita uma base de comparação competente. Abaixo está disposto um exemplo de ruído encontrado nessa base, o que gera desafio para a tarefa de anotação automática.

O que o falante disse: *"generations of **armenians**"*
O que foi transcrito pelo ASR: *"generations of **romanians**"*

Os transcritos foram submetidos às abordagens de anotação na íntegra (sem pré-processamento ou segmentação), assim como estão disponibilizados. Em todas as abordagens de anotação automática utilizadas na experimentação foram usados todos os recursos gerados. Nenhum tipo de limitação de recursos foi imposta nas abordagens para os testes de eficiência. Apenas no *baseline* foi limitado o número de anotações que seriam válidas. Esse limite foi calculado através da frequência da palavra no texto. Eram considerados apenas os recursos correspondentes às palavras que apareciam menos de 20 vezes no texto e mais de 1 vez. Nas demais abordagens todos os recursos foram considerados. Todas as anotações encontradas nesta pesquisa estão ligadas aos recursos da DBpedia.

Se for considerado que muitos vídeos em um repositório não possuem nenhum metadado associado, ou então possuem metadados insignificantes, qualquer informação que esteja presente no conteúdo desse vídeo, ainda que seja genérica, já é válida para identificá-lo, pois muitas vezes não é possível para o usuário assistir a todos os vídeos do repositório para encontrar o que procura.

Um modo rústico e trivial de conseguir anotações para o vídeo pode ser anotar todas as palavras que são ditas no áudio do vídeo. Ainda que simples, essa abordagem pode ser considerada como base para qualquer outra estratégia automática usada para identificar um vídeo por meio de anotação de transcritos. Considerando esse fato, esse foi o ponto de partida da análise de abordagens para a anotação semântica automática e é considerada como técnica *baseline*.

Inicialmente foram anotadas todas as palavras do transcrito que possuíam correspondentes fiéis entre os recursos de uma base de conhecimento, nesse caso foi utilizado a

DBpedia. Porém, essa técnica é muito simplória e cria outras perspectivas, como não simplesmente anotar um vídeo, mas sim gerar uma anotação semântica que permita entender o contexto do vídeo.

Além disso, muitas palavras, quando vistas isoladamente, recebem significados diferentes do que vistas como parte de uma expressão, como é o caso da palavra "rio"que significa "curso de água natural, que deságua em outro rio, no mar ou num lago". Porém, quando vista em conjunto com outras palavras como "Rio Branco"possui o significado de: "município brasileiro, capital do estado do Acre, na Região Norte do país".

Por esse motivo considerou-se também uma abordagem que fosse capaz de criar anotações de nomes compostos, então foi utilizado um Chunker. Após conhecer essas possíveis abordagens, considerou-se outras opções mais elaboradas. Como é o caso do NER que identifica nomes de pessoas, organizações e locais como identificadores viáveis para um documento.

Foi explorado também um modelo de *Entity Linking* para anotações automáticas, com o objetivo de reconhecer e desambiguar textos não estruturados. O modelo de *Entity Linking*, por anotar as palavras do texto individualmente, geram muitos identificadores, podendo encontrar identificadores repetidos. Por esse motivo, analisou-se duas opções, uma aceitando anotações duplicadas e outra aceitando apenas anotações exclusivas (sem duplicatas). E por fim, utilizou-se também um TVSM, que encontra as possíveis interpretações para os termos em um documento, atribuindo maior importância para uma interpretação, quanto mais exclusivos forem seus termos.

As anotações que foram combinadas são: o "NER & CHUNKER", onde foram utilizadas as anotações do NER seguidas pelas anotações do *Chunker*, para cada documento; o "NER || CHUNKER", pois, como o NER gera anotações muito específicas, em alguns casos, o algoritmo não encontra anotações para o documento, nesses casos usou-se o *Chunker* para anotar os documentos que não receberam anotações do NER.

4.2.2 RESULTADOS

A Tabela 4.1 lista todas as abordagens testadas por essa pesquisa para a tarefa de anotação automática e seus respectivos resultados, utilizando as métricas apresentadas na seção anterior:

Na Tabela 4.1 estão dispostos os resultados, onde inicialmente apresenta-se um teste

Experimento	Média Precisão	Média Cobertura	Média TopN
BASELINE	0,0045	0,2878	0,0550
CHUNKER	0,0125	0,2364	0,0603
NER	0,1221	0,0592	0,0707
NER & CHUNKER	0,0160	0,2402	0,0951
NER || CHUNKER	0,1252	0,1160	0,0800
ENTITY LINKING	0,0089	0,4704	0,0883
MODELO DE TÓPICOS	0,0735	0,3420	0,2809

Tabela 4.1: Testes experimentais das abordagens

base neste trabalho, dado a sua simplicidade. Anotar todas as palavras mencionados no texto e encontrar recursos que as identifiquem é o tipo de anotação automática mais simples que pode ser adotado. Então utiliza-se essa anotação como base de comparação para as demais neste trabalho. Esperava-se que técnicas mais complexas que essa trariam melhores resultados, como realmente ocorreu. O fato do *Baseline* alcançar uma cobertura significativa se dá por esse anotar tudo o que é dito no texto, o que não caracteriza-o superior as abordagens com cobertura inferior.

Em seguida estão os resultados do *Chunker*, que não se destacou quando utilizado de forma isolada, porém sua implementação foi relevante visto que unida a outra abordagem gerou resultados consideráveis.

Logo após, estão os resultados do NER, onde tem-se um dos melhores ganhos em precisão de todos os testes feitos, o que indica a relevância dos substantivos em anotações semânticas.

Em seguida estão os resultados dos testes da união das duas técnicas, "NER & CHUNKER", onde unem-se as anotações geradas por ambas as abordagens em um único documento. Essa técnica apresentou o segundo melhor resultado para o topN indicando que as tags válidas que essa abordagem encontra estão melhores posicionadas que nas demais abordagens perdendo apenas para o MODELO DE TÓPICOS.

Logo após, estão os resultados das duas técnicas "NER || CHUNKER", onde foram anotados a maioria dos documentos usando a técnica de NER, contudo onde ela não retornava nada relevante, anotou-se com a técnica do *Chunker*, essa obteve a melhor precisão de todos os testes, por consequência do NER já ser eficiente, ou seja, utilizar o *Chunker* em complemento só agregou valor aos resultados.

No próximo resultado, se encontra os testes feitos utilizando o *Entity Linking* que obteve a melhor cobertura entre todas as apresentadas, pois como mostrado no exemplo

da seção 3.1.3, o *Entity Linking* anota muitos recursos possíveis para um texto. Essa abordagem também apresenta um valor considerável de topN.

Por fim, o resultado do MODELO DE TÓPICOS, técnica que equilibra precisão e cobertura relevantes e o melhor topN.

Como esperado, os resultados mostraram a superioridade dos métodos mais complexos, onde três abordagens se destacam das demais: uma com a melhor precisão NER, outra com a melhor cobertura *Entity Linking* e por fim uma que equilibra boa precisão, boa cobertura e um topN alto, o MODELO DE TÓPICOS.

No Gráfico 4.1 estão todos os resultados obtidos na precisão (linha em vermelho) e na cobertura (linha em azul) separados por abordagens. O gráfico apresenta os vídeos ordenados pela métrica (eixo X) e os valores alcançados pela métrica (eixo Y). Neste gráfico, os vídeos foram ordenados de forma decrescente por métrica para permitir uma visão individual mais acurada dos testes em cada programa. Assim o leitor pode identificar com maior precisão a proporção de programas com resultados bons e ruins em cada experimento.

No Gráfico 4.1 não estão dispostos os resultados do *Baseline*, por esse ter servido apenas como base de comparação, a fim de demonstrar a eficiência das demais abordagens. Verifica-se nos gráficos 4.1(A), 4.1(C), 4.1(E) e 4.1(F) os melhores valores de cobertura. No gráfico 4.1(C) percebe-se uma cobertura muito próxima a do gráfico 4.1(A), isso acontece por característica da abordagem *Chunker* possuir um nível de cobertura relevante é possível verificar ainda na Tabela 4.1 que a cobertura da abordagem "NER & CHUNKER"4.1(C) é um pouco superior a do gráfico 4.1(A), isso por essa abordagem abranger também a cobertura do NER.

Quanto aos gráficos com as melhores valores de precisão estão o 4.1(B) e 4.1(D) assim como no caso da cobertura a abordagem composta "NER || CHUNKER"gráfico 4.1(D), apresenta um resultado próximo ao do gráfico 4.1(B) por carregar uma forte característica de uma das abordagens, nesse caso o NER. Porém é possível observar que gráfico 4.1(D) possui os valores de precisão e cobertura um pouco superiores ao do gráfico 4.1(B) isso acontece pelo fato do *Chunker* agregar na abordagem.

No gráfico 4.1(F) verifica-se valores tanto de precisão quanto de cobertura equilibrados por uma grande gama de programas. O que representa uma abordagem com anotações relevantes para um grande número de programas ao contrário das demais que apresentam

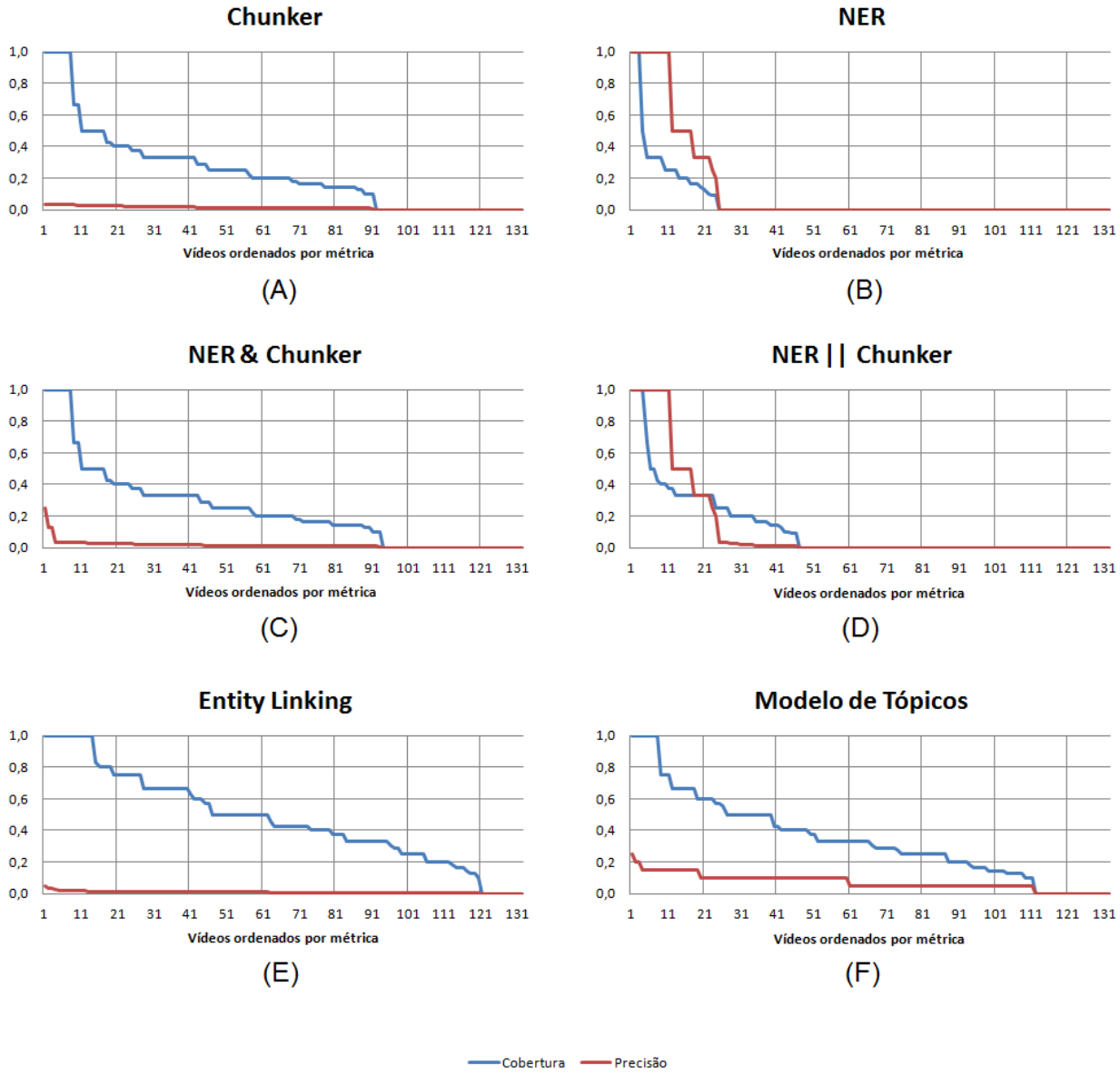

Figura 4.1: Gráfico das abordagens de anotação

valores muito relevantes para poucos programas como é o caso de 4.1(B) e 4.1(D), ou então apresenta uma cobertura boa e níveis de precisão baixos como é o caso de 4.1(A) e 4.1(C) e 4.1(E).

Nota-se em todos os gráficos ocorrências de programas com valores tanto de precisão quanto de cobertura igual a zero, esse fato não indica que os programas não receberam anotações, apenas que as anotações associadas automaticamente não eram relevantes de acordo com os editores, ou seja, todas as anotações geradas são falsos positivos.

O Gráfico 4.2 apresenta a quantidade de abordagens (representada pelas cores) que

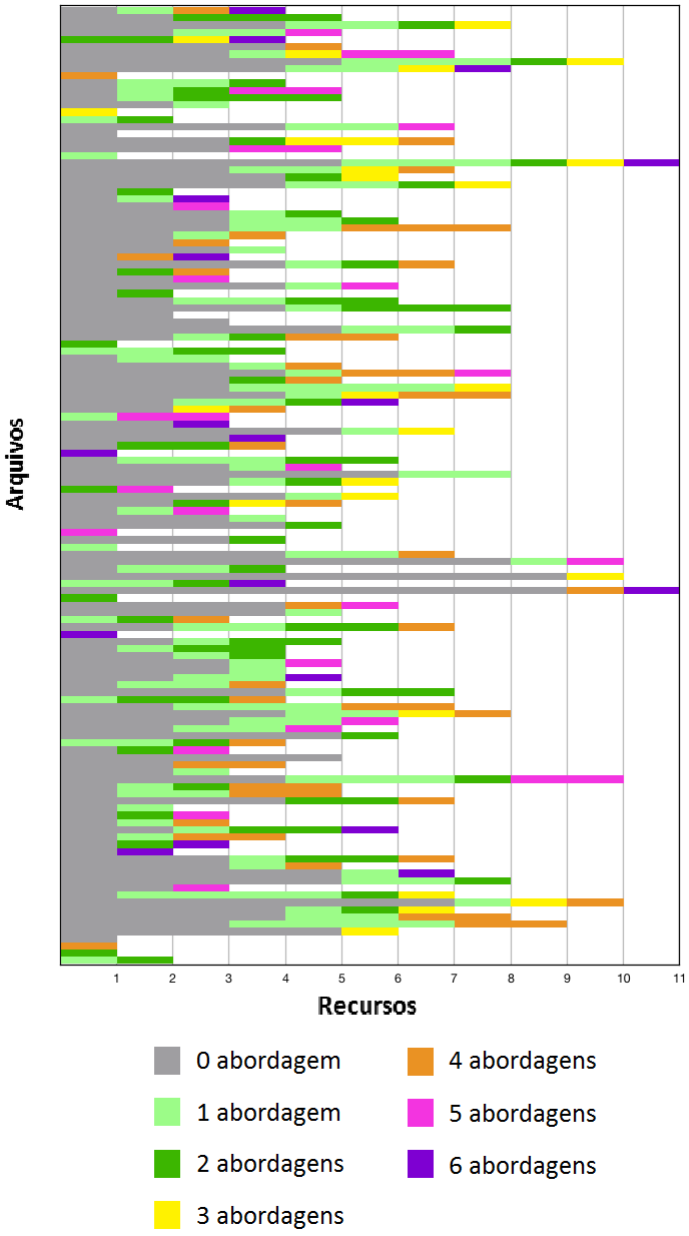

Figura 4.2: Gráfico de quantidade de abordagens para cada tag

anotam uma tag esperada como correta. Verifica-se no gráfico as tags esperadas (eixo X) em relação aos programas (eixo Y). No gráfico é possível observar que muitas tags recebem a cor cinza indicando que nenhuma abordagem conseguiu encontrá-las. Isso acontece pois muitas tags que esperavam ser encontradas não estarem explícitas no áudio, ou seja, recursos que foram inferidos pelos editores profissionais. Os pontos de maior destaque do gráfico são as tags que foram anotadas por poucas abordagens, como uma, duas ou três abordagens. O fato de poucas abordagens reconhecerem a tag, destaca a relevância da abordagem que conseguiu realizar essa tarefa com sucesso. Descobrir quais as abordagens

que conseguem encontrar tags que as demais abordagens não encontram, passa a ser relevante. Considerando essa relevância apresenta-se a Tabela 4.2. Ainda no gráfico 4.2 é possível analisar as cores rosa e roxa representando as tags que são anotadas por 5 ou 6 (todas) as abordagens, essa característica demonstra a relevância das abordagens escolhidas para serem analisadas nesta pesquisa. Essas tags ressalta a capacidade de todas as abordagens anotadoras analisadas reconhecer recursos que os editores também reconhecem.

Abordagens	**Recursos**
Entity Linking	109
Modelo de Tópicos	22
Entity Linking e Modelo de Tópicos	69
Chunker e Ner&Chunker	8
Chunker, Ner&Chunker e Modelo de Tópicos	13
Chunker, Ner&Chunker e Entity Linking	9
Chunker, Ner&Chunker e Ner||Chunker	3

Tabela 4.2: Resultados das Abordagens de Anotação

O Gráfico 4.3 apresenta de forma quantitativa os dados dispostos no Gráfico 4.2, onde no (eixo X) apresenta a quantidade de abordagens que anotam uma tag esperada como certa. Os rótulos sobre as barras e o eixo Y representam a quantidade de tags encontradas pela quantidade de abordagens dispostas no eixo X. Esse gráfico torna possível observar de forma quantitativa as informações apresentadas anteriormente sobre o Gráfico 4.2.

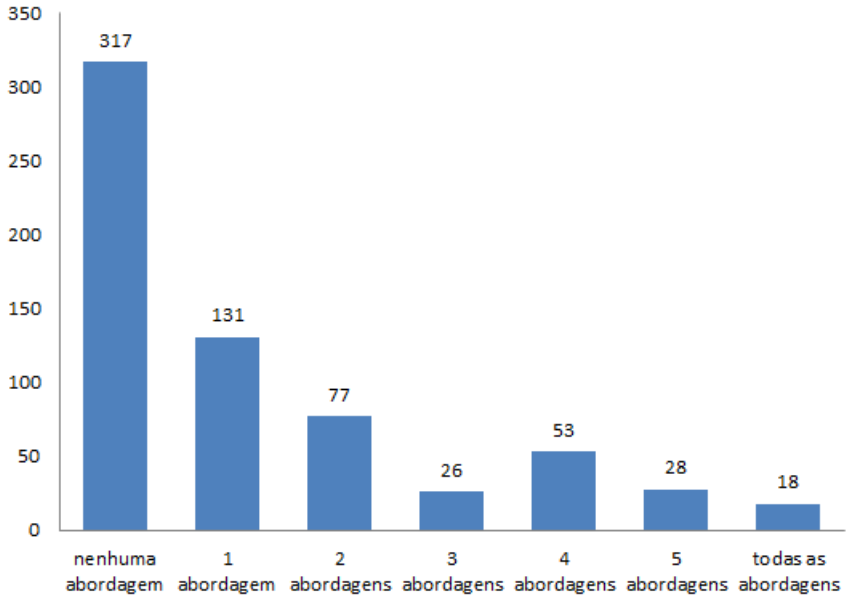

Figura 4.3: Gráfico da quantidade de abordagens anotadoras por quantidade de tag

Na Tabela 4.3 é possível verificar o tempo de processamento por arquivo de cada abordagem, onde observa-se como maior tempo o gasto pela abordagem Modelo de Tópicos, devido ao seu nível de complexidade avançado para realizar a tarefa de anotação, especificamente lidando com os textos ruidosos.

Abordagens	**Média do Tempo de Processamento por Arquivo**
Chunker	2,35s
Ner||Chunker	2,40s
Ner	2,41s
Entity Linking	4,38 s
Ner&Chunker	5s
Modelo de Tópicos	4h 40min

Tabela 4.3: Tempo de Processamento das abordagens

4.3 CÁLCULO DE SIMILARIDADE

Para o cálculo da similaridade os programas foram anotados por cada abordagem individualmente. O algoritmo então usa essas anotações como apresentado na seção 3.2 para encontrar programas que compartilhassem das mesmas categorias, mostrando assim a relação entre eles.

4.3.1 EXPERIMENTOS PARA DEFINIÇÃO DE α E β

O processo do cálculo de similaridade entre vídeos (seção 3.2) utiliza o caminhamento no grafo de categorias da DBpedia para identificar os possíveis relacionamentos. Entretanto, esse caminhamento no grafo pode ser feito em diferentes níveis, caminhando tanto para as categorias mais abrangentes quanto para as mais específicas. Porém, esse nível de profundidade no caminhamento pode influenciar na obtenção de resultados.

4.3.1.1 Criação da base de avaliação

O melhor nível de caminhamento no grafo de categorias da DBpedia foi investigado em (DIAS et al., 2017), onde foi criada uma base de vídeos com relacionamentos manuais definidos por especialistas da área de Ciências Exatas. Nesta base, para cada vídeo, foi informado o relacionamento deste com os demais vídeos existentes no repositório. A

base de avaliação possui 41 vídeos, sendo esses das áreas de Ciência da Computação, Estatística, Química e Física.

A base foi populada com todas as videoaulas do Instituto de Ciências Exatas da UFJF presentes no site VideoAula@RNP no idioma português. Essas aulas foram assistidas por especialistas convidados a realizar duas tarefas. Primeiro, cada especialista atribuiu um recurso da DBpedia para cada assunto explicitamente falado durante o vídeo. Após assistir um conjunto de vídeos, cada especialista informou quais os vídeos mais relacionados a cada vídeo assistido. Não houve restrição do número de recursos e da quantidade de relações para cada vídeo que o especialista poderia atribuir. Ainda, a seleção dos vídeos relacionados seguiu critérios pessoais de cada especialista.

Como a definição dos critérios seguiram critérios pessoais, os relacionamentos finais são a união dos relacionamentos atribuídos por cada especialista, não tendo sido utilizada nenhuma filtragem posterior. Assim, um especialista pode considerar que um vídeo é relacionado por abordar exatamente um mesmo tópico de aula e outro especialista pode relacionar outro vídeo por considera-lo complementar a um pequeno assunto abordado durante a aula. No total foram definidos manualmente 211 relacionamentos, com uma média de 5 relacionamentos por vídeo. Nenhum dos vídeos da base ficou sem relacionamento, apesar de existirem 10 vídeos com apenas 1 relacionamento, 17 vídeos tiveram 5 relacionamentos ou mais.

4.3.1.2 Resultados

Nos experimentos realizados, foram alterados os parâmetros α e β do algoritmo de forma a recuperar diferentes níveis de informações de categorias da DBpedia em cada um dos experimentos. Assim, pode-se analisar como a quantidade de informação processada influencia na acurácia do processo e no tempo de processamento. Os resultados alcançados nas métricas de cobertura e topN em cada experimento estão descritos nas Tabelas 4.4 e 4.5, respectivamente.

$\alpha\backslash\beta$	0	1	2
0	0,67946	$0,91412$	$0,93120$
1	$0,91800$	$0,93120$	$0,93120$
2	$0,93120$	$0,93120$	$0,93120$

Tabela 4.4: Média da cobertura em cada experimento

$\alpha \backslash \beta$	0	1	2
0	0,55103	0,66778	0,66776
1	0,66761	0,67732	0,63420
2	0,61300	0,62910	0,60832

Tabela 4.5: Média do topN em cada experimento

Nos resultados alcançados em (DIAS et al., 2017), verifica-se que caminhar em categorias mais abrangentes (α) e mais específicas (β) produzem maior cobertura. Porém, a cobertura não muda quanto maior for a profundidade. Por outro lado, embora o uso das categorias possa ajudar em um melhor topN, aumentar a profundidade não implica em aumento do topN. Ao aumentar a profundidade, mais categorias em comum dos vídeos terão e, com isso, mais difícil será ranquear corretamente o resultado através do número de categorias em comum. Este efeito é mais evidente ao caminhar com maior profundidade nas categorias mais abrangentes (α=2). A melhor configuração encontrada para o algoritmo foi α=1 e β=1. Então os valores de profundidades adotados nesse trabalho são α=1 e β=1.

4.3.2 EXPERIMENTOS PARA IDENTIFICAÇÃO DA INFLUÊNCIA DA ANOTAÇÃO NO CÁLCULO DE SIMILARIDADE

Para avaliação do cálculo da similaridade entre vídeos foi utilizada a base de vídeos disponibilizada por Raimond and Lowis (2012), com os programas da BBC, os mesmos descritos na seção 4.2.1. A base foi relacionada manualmente para realizar esses experimentos. Não houve restrição do número de relações que seria atribuída para cada vídeo. Nesta base, para cada vídeo foi informado o relacionamento deste com os demais vídeos existentes no repositório. Esperava-se que a abordagem encontrasse novos relacionamentos além dos existentes como descrito em (DIAS et al., 2017). Dos 132 programas 40 vídeos não foram relacionados. A base de relacionamentos foi constituído de 92 vídeos e 94 relacionamentos. Desses, 40 vídeos receberam apenas 1 relacionamento os demais receberam 2 ou mais. O vídeo que recebeu mais relacionamentos obteve 6 relacionamentos.

A linha de base *Baseline* considerada nesse experimento é diferente da utilizada no experimento anterior. Nesse experimento foi considerado como *Baseline* as similaridades encontradas através do uso de anotações manuais, as que foram anotadas pelos próprios editores da BBC.

4.3.2.1 Definição dos métodos de ranqueamento

Como os resultados são ranqueados para cada vídeo, é possível utilizar um método de poda (*threshold*) no resultado para retornar apenas um conjunto limitado de vídeos no conjunto resultante, mas de alta precisão.

Para realizar os experimentos do cálculo de similaridade foram adotados dois métodos de ranqueamento. O primeiro ranqueamento é realizado em relação ao vídeo que está buscando os seus relacionamentos, por exemplo considerando um vídeo A, onde é analisado o quanto o vídeo B pode estar relacionado a A. Então considerando que o vídeo B contém uma porcentagem de categorias satisfatórias de A ele é um vídeo que pode ser relacionado ao vídeo A. Caso nessa mesma investigação, um terceiro vídeo C possível a ser relacionado ao vídeo A não obtenha uma porcentagem satisfatória de A, esse então será podado e não entrará na lista de relacionados do vídeo A. Como valores de porcentagem de similaridade tolerável para a poda, foi considerado 20%. No exemplo, esse valor representa a porcentagem de o quanto de A está em B. Considerou-se também podas de 25% e 30% de similaridade. Essa poda ocorre como a fórmula abaixo.

Método (1):

$$percentagemIntersecao1 = \frac{TotalCategoriaRelacionadas}{numCategoriasDoVideoRef}$$

Onde $percentagemIntersecao1$ representa o valor da porcentagem de categorias relacionadas entre dois vídeos, o que é buscado relacionamento para ele vídeo A e o que é buscado vídeo B. $TotalCategoriaRelacionadas$ representa total de categorias encontradas como relacionadas entre dois vídeos A e B. Por fim, $numCategoriasDoVideoRef$ representa o total de categorias do vídeo que busca-se relacionamento, ou seja, vídeo A.

A segunda forma de ranqueamento ocorre através do cálculo do Coeficiente de Sorensen-Dice, ou seja esse ranqueamento ocorre dando peso dobrado para as co-ocorrências positivas de categorias de B em A, em vez de considerar o quanto de um vídeo A está em B, considera-se o quanto o vídeo B tem de categorias do vídeo A, atribuindo um peso dobrado para elas e dividindo pelo total de categorias de A mais o total de categorias de B. Esse método de ranqueamento pode ser visto na fórmula abaixo.

Método (2):

$$percentagemIntersecao2 = \frac{2*|TotalCategoriaRelacionadas|}{|numCategoriasDoVideoRef|+|TotalCategorias|}$$

Onde $percentagemIntersecao2$ representa o valor da porcentagem de categorias relacionadas entre dois vídeos, o que é buscado relacionamento para ele vídeo A e o que é buscado vídeo B. $TotalCategoriaRelacionadas$ representa total de categorias encontradas como relacionadas entre dois vídeos A e B. Por fim, $numCategoriasDoVideoRef$ representa o total de categorias do vídeo que busca-se relacionamento, ou seja, vídeo A e $TotalCategorias$ o total de categorias do vídeo que está sendo buscado, ou seja, vídeo B.

4.3.2.2 Resultados

As Tabelas 4.6 e a 4.7 apresentam os resultados dos experimentos para cada uma das abordagens de anotação e sua influência para a tarefa de cálculo de similaridade. Na Tabela 4.6 o método de ranqueamento adotado é o primeiro apresentado anteriormente e seus respectivos resultados e na Tabela 4.7 o método é o segundo, através do cálculo do Coeficiente de Sorensen-Dice e seus respectivos resultados.

MÉTODO (1)	MÉDIA PRECISÃO			MÉDIA COBERTURA			MÉDIA TOPN		
EXPERIMENTOS	(20%)	(25%)	(30%)	(20%)	(25%)	(30%)	(20%)	(25%)	(30%)
BASELINE	0,1304	0,1564 ↑	0,1729 ↑	0,5336	0,4948 ↓	0,4460 ↓	0,4284	0,4084 ↓	0,3829 ↓
CHUNKER	0,0210	0,0329 ↑	0,0311 ↓	0,5583	0,4011 ↓	0,2756 ↓	0,1463	0,1457 ↓	0,1341 ↓
NER	0,0118	0,0122 ↑	0,0127 ↑	0,1236	0,1178 ↓	0,1092 ↓	0,0301	0,0296 ↓	0,0327 ↑
NER & CHUNKER	0,0214	0,0289 ↑	0,0264 ↓	0,6006	0,3537 ↓	0,2488 ↓	0,1524	0,1540 ↑	0,1399 ↓
NER || CHUNKER	0,0130	0,0151 ↑	0,0115 ↓	0,1609	0,1465 ↓	0,1322 ↓	0,0533	0,0529 ↓	0,0471 ↓
ENTITY LINKING	0,0202	0,0260 ↑	0,0231 ↓	0,6940	0,4845 ↓	0,2934 ↓	0,1029	0,0987 ↓	0,0831 ↓
MODELO DE TÓPICOS	0,0768	0,1127 ↑	0,0828 ↓	0,4874	0,2776 ↓	0,2029 ↓	0,2632	0,2313 ↓	0,1742 ↓

Tabela 4.6: Testes experimentais das aplicações no cenário de busca por similaridade utilizando o método (1)

MÉTODO (2)	MÉDIA PRECISÃO			MÉDIA COBERTURA			MÉDIA TOPN		
EXPERIMENTOS	(20%)	(25%)	(30%)	(20%)	(25%)	(30%)	(20%)	(25%)	(30%)
BASELINE	0,1534	0,1561 ↑	0,2077 ↑	0,5440	0,4724 ↓	0,4388 ↓	0,4364	0,4051 ↓	0,3909 ↓
CHUNKER	0,0189	0,0233 ↑	0,0507 ↑	0,5426	0,3106 ↓	0,2353 ↓	0,1405	0,1329 ↓	0,1398 ↑
NER	0,0135	0,0148 ↑	0,0143 ↓	0,1322	0,1236 ↓	0,1020 ↓	0,0329	0,0368 ↑	0,0272 ↓
NER & CHUNKER	0,0239	0,0274 ↑	0,0495 ↑	0,5606	0,3451 ↓	0,2331 ↓	0,1552	0,1384 ↓	0,1405 ↑
NER || CHUNKER	0,0139	0,0148 ↑	0,0148 =	0,1466	0,1293 ↓	0,1078 ↓	0,0452	0,0386 ↓	0,0423 ↑
ENTITY LINKING	0,0228	0,0211 ↓	0,0301 ↑	0,6431	0,4414 ↓	0,2029 ↓	0,0980	0,0882 ↓	0,0756 ↓
MODELO DE TÓPICOS	0,0558	0,1148 ↑	0,1287 ↑	0,4816	0,3006 ↓	0,2172 ↓	0,2497	0,2370 ↓	0,1967 ↓

Tabela 4.7: Testes experimentais das aplicações no cenário de busca por similaridade utilizando o método (2)

Na Tabela 4.6 e na Tabela 4.7 estão os valores encontrados a partir dos dois métodos de ranqueamento, divididos em grupos das médias de cada métrica: precisão, cobertura e

topN e em subgrupos de podas de 20%, 25% e 30%. Método (1) e (2) estão relacionados ao método (1) e método (2) respectivamente.

Quando é realizado um corte de todas os programas relacionados, considera-se 20% um corte menor do que 30%. Então os valores das tabelas representam de cortes menores a cortes mais severos. É possível observar em ambas as tabelas que a precisão na maioria das abordagens de anotação aumentaram quando o corte aumentou de 20% para 25%, enquanto quando aumentou um pouco mais, algumas começaram a cair como o *Chunker* na tabela 4.6. Isso ocorre porque a precisão aumenta na medida que o conjunto considerado é restringido, mas quando o conjunto começa a se restringir ainda mais, pode ocorrer de alguns vídeos que eram considerados, não serem mais, por estarem abaixo do limite. É possível observar também que em alguns casos o corte reduz o valor da precisão, como na abordagem *Entity Linking* na tabela 4.7. Porém, um próximo corte sucessivo, consegue aumentar novamente esse valor.

Observa-se também que, quando as podas se tornam maiores, as coberturas de todas as abordagens caem. Tal comportamento é esperado, pois o conjunto está sendo limitado e algumas opções menos relevantes se perdem. Em relação ao topN, a tendência é a mesma das coberturas, ou seja, de terem seus valores reduzidos a medida que a poda aumenta. Porém, em alguns casos onde a precisão cresce, o topN também cresce, como é o caso do "NER & CHUNKER"na tabela 4.7. O topN aumenta com o corte caso não ocorra variação na precisão, como é o caso do "NER||CHUNKER"na tabela 4.7, ou então caso essa precisão também aumente.

O MODELO DE TÓPICOS apresenta os melhores valores em relação aos demais. Analisando a tabela 4.7 com o método (2), seus valores de precisão aumentam em todos os cortes e seus valores de topN são superiores. O MODELO DE TÓPICOS perde apenas para o melhor caso que é o *Baseline*, seguido da abordagem "NER & CHUNKER"que apresenta o segundo comportamento mais satisfatório onde com o método (2) de ranqueamento na tabela 4.7, "NER & CHUNKER"aparece com o segundo melhor topN entre as abordagens.

A seguir está disposto o Gráfico 4.4 e o Gráfico 4.5 com os valores dos topN's (eixo Y) alcançados em cada abordagem (eixo X). O losango azul apresenta os topN's com poda de 30%, o quadrado vermelho os topN's com poda de 25% e os triângulos verdes os topN's com poda de 20%.

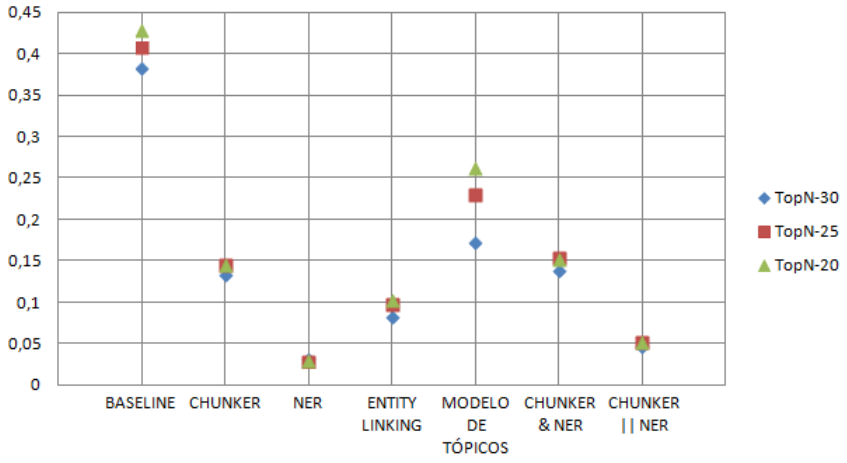

Figura 4.4: Gráfico do topN da similaridade entre vídeos utilizando o método (1)

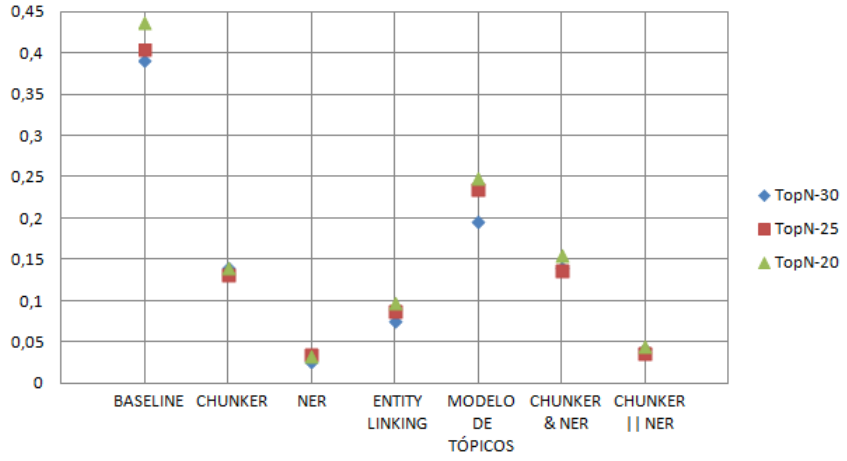

Figura 4.5: Gráfico do topN da similaridade entre vídeos utilizando o método (2)

Nos Gráficos 4.4 e 4.5 é possível visualizar o potencial de topN a ser alcançado através do *Baseline*. A abordagem que alcançou os valores de topN mais relevantes foi o MODELO DE TÓPICOS, seguido da abordagem com o segundo melhor valor, o "NER & CHUNKER". Verifica-se através do gráfico que a poda de 20% apresenta os melhores topN's, isso ocorre devido ao fato de os vídeos não serem ranqueados de forma diferente, o que indica que quantos mais vídeos são admitidos como similares, ou seja, podas menores mais vídeos entram para o cálculo do topN agregando assim valor ao resultado.

No Gráfico 4.6 estão dispostos os valores que as abordagens possuíam na anotação (ponto verde) e os valores que essas atingiram no cálculo de similaridade (ponto vermelho). Neste gráfico possível visualizar o impacto que cada que a precisão e a cobertura das abordagens de anotação exerceram no cálculo da similaridade entre vídeos. Partindo dos pontos verdes que representam o valor das métricas na anotação até os pontos vermelhos

que representam os valores das métricas na similaridade. Para construir tal gráfico foram usados os valores da tabela 4.1 para a anotação e os valores da tabela 4.7 com poda de 30% para o cálculo da similaridade.

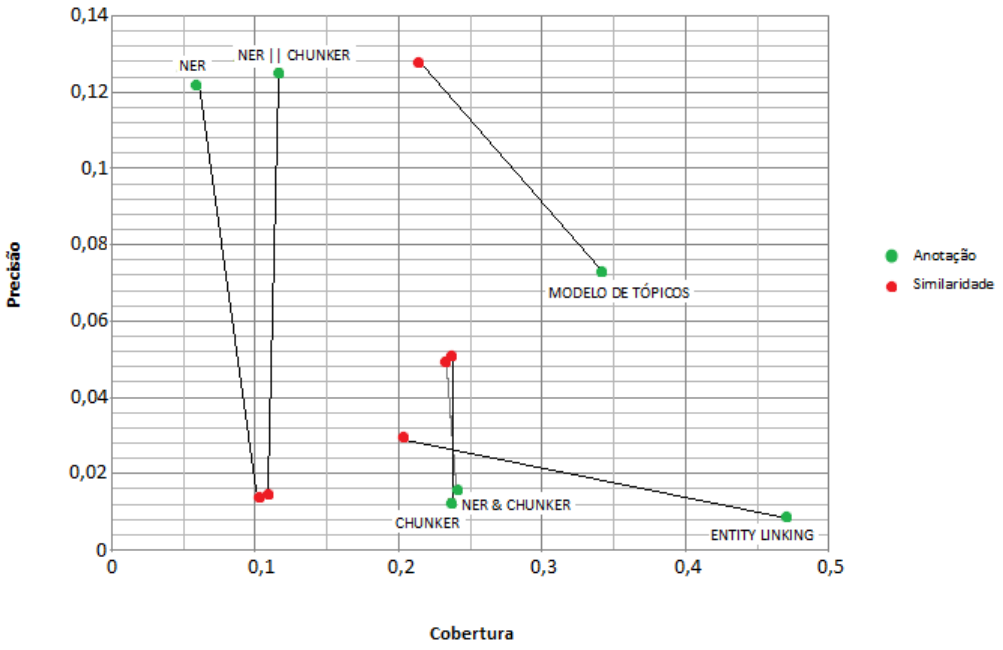

Figura 4.6: Gráfico de correspondências da anotação com o cálculo da similaridade

Pode-se verificar no gráfico 4.6 que a abordagem "NER || CHUNKER"não conseguiu manter sua alta precisão no cenário aplicado. O mesmo ocorreu para o NER. Porém, este último conseguiu atingir um pouco mais de cobertura. Isso pode ser explicado pelo fato dos recursos anotados por essas abordagens não possuírem tantas categorias em comum quanto eram exigidas para se manterem acima dos limites de poda.

O MODELO DE TÓPICOS e o *ENTITY LINKING* perderam em cobertura enquanto ganharam em precisão, no MODELO DE TÓPICOS o ganho em precisão foi mais acentuado, assim como sua perda de cobertura foi menor.

O *Chunker* e o "NER & CHUNKER"ganharam em precisão enquanto obtiveram praticamente a mesma cobertura, no resultado da aplicação essas duas abordagens apesar de não obterem ganho em cobertura, obtiveram os valores de coberturas mais significantes. Tal comportamento, pode ser explicado através do fato da abordagem anotar recursos que compartilhavam muitas categorias, porém, não serem os recursos considerados mais relevantes. O MODELO DE TÓPICOS obteve o melhor resultado em precisão dos experimentos pelo mesmo motivo das abordagens *Chunker* e o "NER & CHUNKER".

Os resultados encontrados através dessa análise mostrou que as melhores abordagens anotadoras para esse cenário são, de acordo com os melhores topN's e as melhores combinações de precisão e cobertura, o MODELO DE TÓPICOS e o "NER & CHUNKER". Porém, é preciso considerar que, apesar da abordagem de MODELO DE TÓPICOS ter uma boa eficiência, essa é muito custosa em questões de tempo de processamento. Em média foi gasto 4 horas e 40 minutos para processar cada arquivo, enquanto o NER & CHUNKER não atinge valores tão altos de eficiência, porém seu custo de processamento é baixo. Em média, foi gasto 5 segundos para processar cada arquivo. Esses valores foram obtidos processando os arquivos em uma máquina com 16GB de memória RAM, 1TB de HD, processador intel core I7-4790 CPU 3.60 GHz, 8 núcleos, sistema operacional Linux Mint 17.3 Cinnamon (64-bit), Kernel do Linux 3.19.0-32-generic e podem ser observados na Tabela 4.3.

5 CONCLUSÕES E TRABALHOS FUTUROS

O aumento de informação disponível em repositórios de conteúdos educacionais é visível e a necessidade de encontrar conteúdo relevante nas buscas, principalmente em vídeos, visto que as informações não estão tão explícitas como nos textos, justifica esta pesquisa.

Neste trabalho foi realizado uma análise de diferentes abordagens para anotação semântica automática em textos ruidosos extraídos por sistemas de ASR. Analisou-se também uma abordagem que utiliza dados ligados para estabelecer relações que não estão explícitas em recursos anotados pelas abordagens de anotação. Além, de investigar o impacto que cada abordagem anotadora causou nessa abordagem de cálculo de similaridade entre vídeos. Para a realização dessas análises considerou-se o domínio de um cenário real o do GT-BAVi, com intuito de abordar nesta pesquisa situações que ocorrem na prática.

A principal base de conhecimento utilizada nessa pesquisa foi a DBpedia. Foi utilizado os relacionamentos descritos em sua ontologia para expandir os recursos anotados nos documentos de um repositório e assim, verificar a similaridade entre diferentes documentos. Desta forma, documentos distintos podem ser identificados como similares mesmo se não possuem as mesmas anotações. Relacionamentos entre estes documentos podem ser feitos automaticamente, usando conceitos que nem sempre um usuário relacionaria.

Como não é trivial optar por uma abordagem de anotação semântica automática dentre as diversas existentes, esse trabalho reuniu um conjunto de 4 abordagens de PLN que realizam essa tarefa e ainda outras duas combinações destas abordagens. De acordo com a análise dos resultados, as abordagens mais eficientes em relação a precisão são a combinação “NER || CHUNKER”, a abordagem NER e o MODELO DE TÓPICOS. Em relação a cobertura as abordagens mais eficientes foram *Entity Linking*, MODELO DE TÓPICOS e “NER & CHUNKER”, isso desconsiderando a abordagem *Baseline*. Em relação ao topN as melhores abordagens foram MODELO DE TÓPICOS, “NER & CHUNKER”e *Entity Linking*.

Confirmando a hipótese desse trabalho, a abordagem mais complexa e que lida com os ruídos o MODELO DE TÓPICOS foi a que obteve melhor eficiência no domínio abordado por esta pesquisa. Como uma opção a essa abordagem, existe ainda uma combinação entre técnica mais comuns “NER & CHUNKER”, essa não é a abordagem mais eficiente, porém

é menos custosa que a mais eficiente e apresenta resultados consideráveis.

Conclui-se então que, dentre todas as abordagens analisadas, essas duas receberam maiores destaques e a opção por uma delas se dá de acordo com a necessidade e os recursos disponíveis para realizar essa tarefa.

5.1 CONTRIBUIÇÕES DO TRABALHO

Como principais contribuições dessa pesquisa estão (1) a comparação entre as principais abordagens utilizadas para anotação automática, (2) o algoritmo de cálculo de similaridade com dois métodos distintos de ranqueamento, (3) a escolha do domínio utilizado para a aplicação das abordagens para anotação automática, e (4) a criação de duas bases de avaliação, uma para as experimentações citadas na seção 4.2[1] e outra para a reprodução das análises desta pesquisa[2]. Bases essas que estão disponíveis para outros pesquisadores replicarem os experimentos e compararem outras abordagens.

5.2 LIMITAÇÕES DO ESTUDO

Este trabalho possui algumas limitações que precisam ser consideradas. Para a análise das abordagens de anotação, esta não é abrangente o suficiente para representar todas as abordagens existentes. Tentou-se, contudo, abranger as abordagens mais frequentes da literatura. Por outro lado, mesmo dentro de uma tarefa de PLN, como o *Entity Linking* ou o Modelo de Tópicos, existem variadas abordagens propostas na literatura que podem gerar resultados distintos. Contudo, embora o resultado possa variar em menor grau, é importante ressaltar que a fraqueza de cada abordagem presente nos experimentos é inerente da tarefa de PLN e não necessariamente da abordagem que foi implementada.

As combinações feitas entre as abordagens também poderiam ser abrangentes a todas as abordagens propostas. Porém, optou-se pelas mais simples devido ao elevado custo de processamento de algumas das abordagens como o Modelo de Tópicos que gasta em média 4 horas e 40 minutos para processar cada arquivo dos programas utilizados na análise.

A escolha do domínio de aplicação das abordagens também foi restrita a uma única opção, o cálculo de similaridade entre vídeos. Entretanto, como observado na seção 2.4 as abordagens de anotação automáticas podem ser usadas para inúmeras finalidades e

[1] https://github.com/ufjf-dcc/LAPIC1-benchmark
[2] https://github.com/ufjf-dcc/LAPIC2-benchmark

em inúmeros cenários distintos. Porém, o que justificou a escolha da finalidade abordada nesse trabalho foi a demanda desse estudo em um cenário real, o qual utiliza o cálculo de similaridade entre vídeos para gerar recomendações a outros vídeos dentro de um repositório de videoaulas da RNP.

5.3 TRABALHOS FUTUROS

Para trabalhos futuros, é interessante analisar a abrangência de outras bases de conhecimento para a busca da similaridade entre vídeos. Possibilitando assim encontrar recursos distintos mais específicos caso seja usada uma base especialista. Contudo, nesse trabalho foi usada a DBpedia, uma base generalista, com o intuito de não excluir nenhum assunto possível de ser abordado nos programas.

Também pretende-se estudar uma forma de tratar o uso de diferentes bases de conhecimento simultaneamente. Ainda, o uso simultâneo de diversas bases de conhecimento gera o problema técnico de processamento desse volume de dados, sendo necessário melhorar a abordagem para realizar filtros durante o processo de busca das categorias.

É possível incluir a abordagem do cálculo de similaridade entre vídeos e os métodos de ranqueamento em um sistema de recomendação de vídeo, que leva em consideração o perfil do usuário. Possibilitando assim recomendar vídeos pelos assuntos abordados neles, como também pelas experiências dos usuários que utilizarem o sistema. Assim, será possível a previsão das recomendações com bases em experiências passadas que são armazenadas pelo sistema.

Pretende-se também realizar limitações de categorias possíveis de serem encontradas para um vídeo na base de conhecimento, impedindo assim que recursos que estão fora do domínio sejam anotados. O que possibilitaria a análise de quais categorias são importantes para o domínio de videoaulas.

Além disso, realizar novas combinações de técnicas que não foram realizadas nessa pesquisa, até mesmo a possibilidade de adotar outras abordagens de anotação que não foram estudadas aqui, como clusterização ou abordagens focadas na fonética das palavras.

Ainda, espera-se adotar outro cenário de aplicação para documentos anotados automaticamente, a fim de identificar se outras características das abordagens anotadoras atendem melhor à cenários distintos.

Por fim, deseja-se considerar outros meios de melhorar o desempenho de abordagens

muito custosas como é o caso do MODELO DE TÓPICOS.

REFERÊNCIAS

ABNEY, S. Parsing by chunks. **Principle-based parsing**, v. 44, p. 257–278, 1991.

AMARAL, D. O. F. d. **O reconhecimento de entidades nomeadas por meio de conditional random fields para a língua portuguesa**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2013.

AQUINO, M. C. Hipertexto 2.0, folksonomia e memória coletiva: um estudo das tags na organização da web. **Revista E-compós**, v. 9, 2007.

ARAÚJO, L. d. R.; SOUZA, J. F. de. Aumentando a transparência do governo por meio da transformação de dados governamentais abertos em dados ligados. **Revista Eletrônica de Sistemas de Informação**, v. 10, n. 1, 2011.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern mation Retrieval: the concepts and technology behind search second edition**, 2011.

BARRÉRE, E. Videoaulas: aspectos técnicos, pedagógicos, aplicações e bricolagem. **Jornada de Atualização em Informática na Educação**, v. 3, n. 1, 2014.

BECKER, J.; KUROPKA, D. Topic-based vector space model. In: **Proceedings of the 6th international conference on business information systems**, 2003. p. 7–12.

BENZEGHIBA, M.; MORI, R. D.; DEROO, O.; DUPONT, S.; ERBES, T.; JOUVET, D.; FISSORE, L.; LAFACE, P.; MERTINS, A.; RIS, C. et al. Automatic speech recognition and speech variability: A review. **Speech communication**, Elsevier, v. 49, n. 10, p. 763–786, 2007.

BERENZWEIG, A.; LOGAN, B.; ELLIS, D. P.; WHITMAN, B. A large-scale evaluation of acoustic and subjective music-similarity measures. **Computer Music Journal**, MIT Press, v. 28, n. 2, p. 63–76, 2004.

CHEN, Z.; CAO, J.; SONG, Y.; ZHANG, Y.; LI, J. Web video categorization based on wikipedia categories and content-duplicated open resources. In: ACM. **Proceedings of the 18th ACM international conference on Multimedia**, 2010. p. 1107–1110.

CHENIKI, N.; BELKHIR, A.; SAM, Y.; MESSAI, N. Lods: A linked open data based similarity measure. In: IEEE. **Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016 IEEE 25th International Conference on**, 2016. p. 229–234.

CHOPRA, A.; PRASHAR, A.; SAIN, C. Natural language processing. **Int. J. Technol. Enhanc. Emerg. Eng. Res**, Citeseer, v. 1, n. 4, p. 131–134, 2013.

CHOWDHURY, G. G. Natural language processing. **Annual review of information science and technology**, Wiley Online Library, v. 37, n. 1, p. 51–89, 2003.

CORRÊA, D. d. L. V. A interpretação semântica de textos científicos em português na perspectiva da ciência da informação: Procedimentos e aplicação à área de ciências agrárias. Universidade Federal de Pernambuco, 2016.

CROFT, W. B.; METZLER, D.; STROHMANN, T. **Search engines**, 2010.

DAHL, G. E. **Deep learning approaches to problems in speech recognition, computational chemistry, and natural language text processing**. Tese (Doutorado) — University of Toronto, 2015.

DASIOPOULOU, S.; GIANNAKIDOU, E.; LITOS, G.; MALASIOTI, P.; KOMPATSIARIS, Y. A survey of semantic image and video annotation tools. In: **Knowledge-driven multimedia information extraction and ontology evolution**, 2011. p. 196–239.

DIAS, L. L.; BARBOSA, J. S.; BARRÉRE, E.; SOUZA, J. F. D. Uma abordagem para identificação de similaridade entre recursos educacionais utilizando bases de conhecimento externas. In: , 2017. v. 25, n. 2.

DURAN, M. S. A importância dos recursos lexicais para o processamento automático do português. **Estudos Linguísticos (São Paulo. 1978)**, v. 42, n. 2, p. 866–877, 2016.

DURAN, M. S.; MARTINS, J. P.; ALUÍSIO, S. M. Um repositório de verbos para a anotação de papéis semânticos disponível na web. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**, 2013. p. 168–172.

ELSAYED, E.; ELKORANY, A.; HAFNY, H. et al. Review aspects of using social annotation for enhancing search engine performance. **International Journal of Computer (IJC)**, v. 24, n. 1, p. 177–189, 2017.

ELSAYED, E.; ELKORANY, A.; SALAH, A.; HAFNY, H. Semantic information retrieval based on social annotation. **International Journal of Computer Science and Information Security**, LJS Publishing, v. 15, n. 3, p. 121, 2017.

FELDMAN, S. Nlp meets the jabberwocky: Natural language processing in information retrieval. **ONLINE-WESTON THEN WILTON-**, ONLINE INC, v. 23, p. 62–73, 1999.

FIGUEIREDO, A. P. S.; ASSIREU, A. T. A. T.; SOUZA, V. C. O. Material didático multimídia aplicado a educação: um relato de experiência na graduação. **Revista Brasileira de Informática na Educação**, v. 23, n. 2, p. 128–136, 2015.

FIGUEIREDO, M. A. Z. d. **MÉTODO PARA REPRESENTAÇÃO DE CONCEITOS POR MEIO DE TÉCNICAS DE ANÁLISE DE TEXTOS EM SEQUENCIA TEMPORAL**. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2017.

FRANK, E.; PAYNTER, G. W.; WITTEN, I. H.; GUTWIN, C.; NEVILL-MANNING, C. G. Domain-specific keyphrase extraction. In: MORGAN KAUFMANN PUBLISHERS INC., SAN FRANCISCO, CA, USA. **16th International Joint Conference on Artificial Intelligence (IJCAI 99)**, 1999. v. 2, p. 668–673.

GRAVIER, G.; JONES, G. F.; LARSON, M.; ORDELMAN, R. Overview of the 2015 workshop on speech, language and audio in multimedia. In: ACM. **Proceedings of the 23rd ACM international conference on Multimedia**, 2015. p. 1347–1348.

GRÜNEWALD, F.; MEINEL, C. Implementation and evaluation of digital e-lecture annotation in learning groups to foster active learning. **IEEE Transactions on Learning Technologies**, IEEE, v. 8, n. 3, p. 286–298, 2015.

GUPTA, Y.; SAINI, A.; SAXENA, A. A new fuzzy logic based ranking function for efficient information retrieval system. **Expert Systems with Applications**, Elsevier, v. 42, n. 3, p. 1223–1234, 2015.

HERRERA, J. E. T.; CASANOVA, M. A.; NUNES, B. P.; LOPES, G. R.; LEME, L. Dbpedia profiler tool: Profiling the connectivity of entity pairs in dbpedia. In: **Proceed-**

ings of the 5th International Workshop on Intelligent Exploration of Semantic Data (IESD 2016).[GS Search], 2016.

HOAD, T. C.; ZOBEL, J. Methods for identifying versioned and plagiarized documents. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 54, n. 3, p. 203–215, 2003.

HULPUS, I.; HAYES, C.; KARNSTEDT, M.; GREENE, D. Unsupervised graph-based topic labelling using dbpedia. In: ACM. **Proceedings of the sixth ACM international conference on Web search and data mining**, 2013. p. 465–474.

JAIN, P.; HITZLER, P.; SHETH, A. P.; VERMA, K.; YEH, P. Z. Ontology alignment for linked open data. In: SPRINGER. **International Semantic Web Conference**, 2010. p. 402–417.

JIANG, Y.-G.; BHATTACHARYA, S.; CHANG, S.-F.; SHAH, M. High-level event recognition in unconstrained videos. **International journal of multimedia information retrieval**, Springer, v. 2, n. 2, p. 73–101, 2013.

KAVASIDIS, I.; PALAZZO, S.; SALVO, R. D.; GIORDANO, D.; SPAMPINATO, C. An innovative web-based collaborative platform for video annotation. **Multimedia Tools and Applications**, Springer, v. 70, n. 1, p. 413–432, 2014.

KAWASE, R.; SIEHNDEL, P.; NUNES, B. P.; HERDER, E.; NEJDL, W. Exploiting the wisdom of the crowds for characterizing and connecting heterogeneous resources. In: ACM. **Proceedings of the 25th ACM conference on Hypertext and social media**, 2014. p. 56–65.

KÖHNCKE, B.; BALKE, W.-T. Using wikipedia categories for compact representations of chemical documents. In: ACM. **Proceedings of the 19th ACM international conference on Information and knowledge management**, 2010. p. 1809–1812.

KORENIUS, T.; LAURIKKALA, J.; JÄRVELIN, K.; JUHOLA, M. Stemming and lemmatization in the clustering of finnish text documents. In: ACM. **Proceedings of the thirteenth ACM international conference on Information and knowledge management**, 2004. p. 625–633.

KÜÇÜK, D.; YAZICI, A. A semi-automatic text-based semantic video annotation system for turkish facilitating multilingual retrieval. **Expert Systems with Applications**, Elsevier, v. 40, n. 9, p. 3398–3411, 2013.

LEHMANN, J.; ISELE, R.; JAKOB, M.; JENTZSCH, A.; KONTOKOSTAS, D.; MENDES, P. N.; HELLMANN, S.; MORSEY, M.; KLEEF, P. V.; AUER, S. et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. **Semantic Web**, IOS Press, v. 6, n. 2, p. 167–195, 2015.

LIDDY, E. D. Enhanced text retrieval using natural language processing. **Bulletin of the Association for Information Science and Technology**, Wiley Online Library, v. 24, n. 4, p. 14–16, 1998.

MAKHOUL, J.; KUBALA, F.; LEEK, T.; LIU, D.; NGUYEN, L.; SCHWARTZ, R.; SRIVASTAVA, A. Speech and language technologies for audio indexing and retrieval. **Proceedings of the IEEE**, IEEE, v. 88, n. 8, p. 1338–1353, 2000.

MEDEIROS, S. F. L.; PANSANATO, L. T. E. Estudo das preferências de alunos e professores sobre videoaula para identificar requisitos de software para ferramentas de produção. **XXVI Simpósio Brasileiro de Informática na Educação**, 2015.

MEINEDO, H. D. dos S. **Audio Pre-processing and Speech Recognition for Broadcast News**. Tese (Doutorado) — Universidade Técnica de Lisboa, 2008.

MENDES, P. N.; JAKOB, M.; GARCÍA-SILVA, A.; BIZER, C. Dbpedia spotlight: shedding light on the web of documents. In: ACM. **Proceedings of the 7th international conference on semantic systems**, 2011. p. 1–8.

NASH, S. Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses. **Interdisciplinary Journal of E-Learning and Learning Objects**, Informing Science Institute, v. 1, n. 1, p. 217–228, 2005.

OLIVEIRA, A. L.; SILVA, E. S.; MACEDO, H. T.; MATOS, L. N. Brazilian portuguese speech-driven answering system. In: ACM. **Proceedings of the 6th Euro American Conference on Telematics and mation Systems**, 2012. p. 277–284.

OLIVEIRA, F. K.; SANTANA, J. R.; PONTES, M. G. O. O vídeo como ferramenta educacional a partir de múltiplas plataformas. **XXI Simpósio Brasileiro de Informática na Educação**, 2010.

OLIVEIRA, P.; ROCHA, J. Semantic annotation tools survey. In: IEEE. **Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on**, 2013. p. 301–307.

OREN, E.; DELBRU, R.; CATASTA, M.; CYGANIAK, R.; STENZHORN, H.; TUMMARELLO, G. Sindice. com: a document-oriented lookup index for open linked data. **International Journal of Metadata, Semantics and Ontologies**, Inderscience Publishers, v. 3, n. 1, p. 37–52, 2008.

PEREIRA, J. W.; GONÇALVES, M. R. B.; SANTOS, M. T. P. Semantic annotation in historical documents. In: IEEE. **Information Systems and Technologies (CISTI), 2017 12th Iberian Conference on**, 2017. p. 1–7.

POLYVYANYY, A. **Evaluation of a novel mation retrieval model: eTVSM**. Dissertação (Mestrado) — Hasso Plattner Institut, February 2007.

QAZI, A.; GOUDAR, R. Emerging trends in reducing semantic gap towards multimedia access: A comprehensive survey. **Indian Journal of Science and Technology**, v. 9, n. 30, 2016.

RAIMOND, Y.; LOWIS, C. Automated interlinking of speech radio archives. **LDOW**, v. 937, 2012.

RAMALHO, J. C. Anotação estrutural de documentos e sua semântica: especificação da sintaxe, semântica e estilo para documentos. 2000.

SACK, H.; WAITELONIS, J. Exploratory semantic video search with yovisto. In: IEEE. **Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on**, 2010. p. 446–447.

SHELKE, N.; DESHPANDE, S.; THAKARE, V. Domain independent approach for aspect oriented sentiment analysis for product reviews. In: SPRINGER. **Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications**, 2017. p. 651–659.

SILVEIRA, M.; NOGUEIRA, V. B.; RODRIGUES, I. Ferramentas e tecnologias para a integração e extração de informação hospitalar. Escola de Ciências e Tecnologia da Universidade de Évora, 2015.

SLIMANI, T. Semantic annotation: The mainstay of semantic web. **arXiv preprint arXiv:1312.4794**, 2013.

SOUZA, J. F. D.; JUNIOR, J. G.; BARRÉRE, E. Comparativo entre fontes de dados para anotação automática de videoaulas. 2017.

TASKIRAN, C. M.; PIZLO, Z.; AMIR, A.; PONCELEON, D.; DELP, E. J. Automated video program summarization using speech transcripts. **IEEE Transactions on Multimedia**, IEEE, v. 8, n. 4, p. 775–791, 2006.

TU, K.; MENG, M.; LEE, M. W.; CHOE, T. E.; ZHU, S.-C. Joint video and text parsing for understanding events and answering queries. **IEEE MultiMedia**, IEEE, v. 21, n. 2, p. 42–70, 2014.

TURNBULL, D.; BARRINGTON, L.; TORRES, D.; LANCKRIET, G. Semantic annotation and retrieval of music and sound effects. **IEEE Transactions on Audio, Speech, and Language Processing**, IEEE, v. 16, n. 2, p. 467–476, 2008.

VINCIARELLI, A. Noisy text categorization. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 27, n. 12, p. 1882–1895, 2005.

VOLKMER, T.; SMITH, J. R.; NATSEV, A. P. A web-based system for collaborative annotation of large image and video collections: an evaluation and user study. In: ACM. **Proceedings of the 13th annual ACM international conference on Multimedia**, 2005. p. 892–901.

WAITELONIS, J.; SACK, H.; HERCHER, J.; KRAMER, Z. Semantically enabled exploratory video search. In: ACM. **Proceedings of the 3rd international semantic search workshop**, 2010. p. 8.

WU, X.; ZHANG, L.; YU, Y. Exploring social annotations for the semantic web. In: ACM. **Proceedings of the 15th international conference on World Wide Web**, 2006. p. 417–426.

XIE, H.; LIU, A.; WANG, F. L.; WONG, T.-L.; LIU, X.; RAO, Y. Revisit tag-based profiles in the folksonomy: How many tags are sufficient for profiling? In: IEEE. **Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on**, 2017. p. 274–277.

YANG, H.; MEINEL, C. Content based lecture video retrieval using speech and video text information. **IEEE Transactions on Learning Technologies**, IEEE, v. 7, n. 2, p. 142–154, 2014.

YE, X. Identify the semantic meaning of service rules with natural language processing. In: IEEE. **Parallel and Distributed Computing, Applications and Technologies (PDCAT), 2016 17th International Conference on**, 2016. p. 63–68.

ZHAO, B.; XU, S.; LIN, S.; LUO, X.; DUAN, L. A new visual navigation system for exploring biomedical open educational resource (oer) videos. **Journal of the American Medical matics Association**, Oxford University Press, v. 23, n. e1, p. e34–e41, 2015.

ZHU, G.; IGLESIAS, C. A. Computing semantic similarity of concepts in knowledge graphs. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 29, n. 1, p. 72–85, 2017.