

Aluno: Airton Bordin Junior

Revisão Sistemática sobre técnicas de criação e expansão automatizada de Dicionários Léxicos no contexto de Análise de Sentimentos

1 Justificativa

A Revisão Sistemática da Literatura fornece uma forma estruturada, objetiva e reprodutível de identificar, avaliar e interpretar trabalhos relevantes em uma determinada área de conhecimento. A análise desses trabalhos apoiará a resolução das questões de pesquisa, que devem ser respondidas pelo presente projeto.

Ao documentar a estratégia e planejamento da revisão, a pesquisa torna-se mais confiável e imparcial, dificultando a seleção somente de artigos que corroboram à tese do pesquisador.

Portanto, de forma a identificar os trabalhos primários e o estado da arte sobre o tema do presente projeto, bem como a seleção e avaliação imparcial dos artigos, justifica-se a utilização do processo de Revisão Sistemática. Além disso, será possível responder as perguntas da pesquisa e identificar *gaps* na área de estudo.

2 Perguntas

- Quais as principais técnicas empregadas na criação e expansão de Dicionários Léxicos?
- Estratégias Evolutivas estão sendo utilizadas no contexto de criação e expansão de Dicionários Léxicos?
- Como a diferença de contexto e domínio vem sendo tratada no contexto de Análise de Sentimentos?
- Que métricas estão sendo utilizadas para avaliar a qualidade dos algoritmos de criação e expansão de Dicionários Léxicos?

3 Palavras chave

- Sentiment Analysis
- Opinion Mining
- Lexicon Expansion
- Genetic Algorithms
- Genetic Programming
- Semantic Orientation

4 Checklist

1. Quais são os objetivos da revisão?

Encontrar, nas principais bases de dados do tema, os trabalhos primários relacionados, de forma a identificar o estado da arte e as abordagens mais comuns no contexto da presente pesquisa.

2. Quais fontes de pesquisa foram utilizadas para identificar os estudos primários?

Para a identificação dos estudos primários, buscas serão realizadas nas bases da *Science Direct*, *IEEEExplore*, *ACM Digital Library*, *Research Gate*, *Semantic Scholar*

3. As fontes utilizadas apresentam algum tipo de restrição?

Alguns artigos não estão disponíveis de forma integral e gratuita.

4. Quais são os critérios de inclusão e exclusão e como estes serão aplicados?

Serão realizadas buscas utilizando as palavras-chave, de forma combinada, nas bases supracitadas, de forma a encontrar artigos primários relevantes sobre o tema. Serão selecionados os artigos publicados à partir do ano 2000 e que estejam disponíveis na íntegra para acesso. Caso a base de dados permita, será feita uma ordenação por relevância, baseada na quantidade de citações e importância do trabalho. Além disso, serão considerados somente artigos escritos em inglês.

Feito isso, será feita a leitura dos títulos e, posteriormente, dos resumos de cada trabalho, de forma a aceitar ou rejeitar o trabalho para a próxima fase da Revisão Sistemática. Os trabalhos rejeitados não serão levados em consideração para o desenvolvimento do projeto, mas ainda assim serão utilizados como dados estatísticos.

Para auxiliar nesse processo, será utilizada a ferramenta StArt (*State of the Art through Systematic Reviews*), desenvolvida pela Universidade Federal de São Carlos, que fornece funcionalidades para apoiar o pesquisador em todas as fases da Revisão Sistemática da Literatura.

5. Quais foram os critérios utilizados para assegurar a qualidade dos estudos primários?

O principal critério utilizado para assegurar a qualidade dos estudos primários é a quantidade de citações que o mesmo possui. Via de regra, um trabalho com muitas citações na literatura pode ser considerado um trabalho de qualidade, uma vez que é utilizado como referência por vários outros.

Outro critério utilizado é uma avaliação quanto à aderência do trabalho às questões de pesquisa definidas. Uma escala de 0 a 5 foi definida, sendo 0 considerado não aderente e 5 completamente aderente. Como há mais de uma questão de pesquisa definida, foi realizada uma média aritmética das escalas.

Importante salientar que esta análise foi realizada por apenas um pesquisador e pode acarretar em uma baixa segurança em relação à validade das avaliações feitas.

6. Como os critérios de qualidade serão aplicados?

O critério relacionado à quantidade de citações será feito por meio de ferramentas disponibilizadas pelas próprias bases de dados, que permitem a ordenação dos trabalhos por quantidade de citações.

A avaliação da escala de aderência do trabalho é feita por meio da leitura do resumo. Caso o mesmo não seja conclusivo ou suficiente para a atribuição da nota, é feita a leitura da introdução e conclusão.

7. Como foram extraídos os dados dos estudos primários?

8. Como os dados serão sintetizados?

Os dados serão tabulados de forma a facilitar a visualização e comparação das abordagens utilizadas nos trabalhos primários consultados. Além disso, as informações mais relevantes, como técnicas utilizadas pela maior parte dos trabalhos, serão apresentadas de forma a demonstrar sua importância, descrevendo detalhes importantes caso mostre-se necessário.

9. Como os dados serão combinados?

Dados que tratam de abordagens semelhantes serão tabulados de forma conjunta, de forma a facilitar a visualização e comparação das informações.

Análises comparativas poderão ser realizadas, demonstrando as principais técnicas e os melhores resultados em determinados contextos.

Análises quantitativas dos processos descritos nos trabalhos primários também poderão ser realizadas, de forma a identificar possíveis pontos que necessitam de estudos para melhoria dos resultados, bem como para apresentar o estado da arte no domínio da pesquisa.

Referências

[Kitchenham and Charters, 2007] Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, Technical report, EBSE Technical Report EBSE-2007-01.