

Desambiguação de nomes de autores usando sistemas de organização de conhecimento disponíveis na Web

JUCIANDER LUIZ MOREIRA (Autor), ANDERSON ALMEIDA FERREIRA (DECOM) (Orientador)

O número de trabalhos publicados em bibliotecas digitais cresce a cada dia. Porém, a identificação do correto autor desses trabalhos não é uma tarefa trivial e é bastante custosa. Além disso, métodos automáticos para fazer tal identificação não resolvem completamente este problema, conhecido como ambiguidade de nomes. Este problema prejudica tanto a recuperação de informação nessas bibliotecas, quanto o seu conceito perante os usuários. A ambiguidade de nomes ocorre devido ao fato de vários nomes poderem referenciar a mesma pessoa, como também pelo fato de várias pessoas terem o mesmo nome. Os métodos automáticos não realizam corretamente a atribuição de todos os nomes aos seus respectivos autores principalmente devido ao fato que, em muitos casos, a escassez de informação e aos dados incompletos presentes nos registros das publicações. Para tentar melhorar os resultados já alcançados pelos métodos de desambiguação, este trabalho propõe o uso da expansão de termos do título das publicações, acrescentando ou substituindo sinônimos ou hiperônimos encontrados em sistemas de organização de conhecimento disponíveis na web. Neste trabalho, inicialmente, foi feito um estudo sobre diversos métodos representativos de desambiguação, selecionando seis métodos, entre métodos supervisionados e não supervisionados, e uma avaliação de diversas bases de conhecimento (SKOS, WordNet e Wikipedia), visando entender a forma de acessar e extrair informações dessas bases. Após ter os novos registros com os títulos alterados, estes foram submetidos aos métodos e os resultados avaliados estatisticamente. Como resultado, existiram algumas situações em que houve pequena melhora, mas em grande parte dos resultados houve empate estatístico. Uma possível explicação é o fato de não terem sido usadas bases de conhecimento específicas da área das publicações. Como trabalho futuro, há a intenção de tentar encontrar bases de conhecimento específicas e avaliar o resultado.

Instituição de Ensino: Universidade Federal de Ouro Preto