

PROTOTIPAGEM DA CLASSIFICAÇÃO AUTOMÁTICA DOS PAPÉIS DE TRECHOS DAS SENTENÇAS PARA PROSPECÇÃO TECNOLÓGICA A PARTIR DE UM ARTIGO TÉCNICO

LIMA, Herlon Augusto Aguiar¹ (herlon214@gmail.com); **BATISTA JUNIOR, Joinvile**² (joinvile@ufgd.edu.br);

¹ Discente do curso de Sistemas de Informação; PIVIC/UFGD;

² Docente do curso de Sistemas de Informação; UFGD;

A proposta original deste trabalho previa a seleção automática de relações extraídas pelo extrator de informação OLLIE, com o objetivo de classificar o papel desempenhado por tais relações no contexto de artigos técnicos. No entanto, após realizada uma análise nas relações extraídas pelo OLLIE, a partir de um artigo técnico escolhido ficou caracterizada a inviabilidade da realização da seleção, em função: (a) dos erros gerados nas extrações de relações resultando em relações extraídas mal construídas e na omissão da extração de muitas relações relevantes; e (b) da perda de contexto de relações em relação à semântica da sentença como um todo. A melhor opção encontrada foi a definição de padrões para classificação automática do papel de trechos relevantes das sentenças no contexto de um artigo técnico. O título do trabalho foi atualizado para melhor expressar os resultados do trabalho realizado. A reorientação do trabalho, não altera o objetivo central da proposta inicial: classificar o papel, no contexto de um artigo técnico, dos trechos relevantes das sentenças que compõem o artigo. A diferença básica é que os trechos relevantes passam a ser identificados diretamente através de padrões e não a partir das relações extraídas por um extrator de informações. A metodologia utilizada incorpora: (1) análise das extrações realizadas pelo extrator de informação OLLIE; (2) especificação de padrões para a classificação do papel das sentenças no contexto do artigo técnico; e (3) implementação e teste de uma ferramenta para aplicar os padrões de classificação a sentenças do artigo técnico. Como resultado deste trabalho: (a) foi especificado um conjunto de padrões de classificação do papel das sentenças em um artigo técnico; e (b) foi construída uma ferramenta para aplicação automática dos padrões às sentenças. Para cada sentença do artigo técnico selecionado, foram especificados, no formato XML, padrões para classificar trechos da sentença. Foi definida um sintaxe para representar os padrões visando futuras generalizações dos padrões, para que os mesmos possam ser utilizados para classificar sentenças de vários artigos técnicos. A sintaxe utilizada para a definição dos padrões é baseada em elementos para representar: (a) trechos de texto contido na sentença original; (b) parâmetros que podem assumir um dos valores de um dos *strings* de uma dado conjunto; (c) trecho do padrão com repetição de 0 a n vezes; (d) trecho opcional do padrão; e (e) abrangência de 1 a n palavras da sentença, com um trecho não específico. A ferramenta para aplicação dos padrões às sentenças, foi utilizada para validar a sintaxe da cada padrão definido. Todas as sentenças do artigo selecionado foram testadas em relação aos padrões derivados a partir de tais sentenças. Este resultado é o passo inicial para generalização dos padrões para classificar sentenças de vários artigos técnicos.

Palavra-chave: Automação de Ontologias. Padrões de Classificação. Classificação de Sentenças.