

CONVERSÃO AUTOMÁTICA, PARA RELAÇÕES BINÁRIAS, DE RELAÇÕES DE ASSOCIAÇÃO EXTRAÍDAS DE FORMA AUTOMATIZADA

Rafael Biassio (rafaelobiassio@gmail.com)

Joinvile Batista Junior (joinvile.batista@gmail.com)

Ontologias são construídas a partir de relações binárias (argumento 1 -- relação -- argumento 2). Os argumentos são majoritariamente associados a uma frase nominal (que não possui verbos e que tem substantivos como foco de sua semântica) e a relação é associada a uma frase verbal (que tem verbos como foco de sua semântica). As extrações geradas pelo extrator aberto disponível OpenIE 4, a partir de sentenças mais complexas, muitas vezes contém frases verbais fora da relação: (a) no argumento 1; (b) no argumento 2; ou (c) em trechos que foram inadequadamente descartados nas extrações geradas. O objetivo deste trabalho é a remoção de frases verbais de argumentos e de trechos descartados, os quais estão sendo referenciados como argumentos complexos e trechos descartados complexos. A metodologia utilizada neste trabalho foi baseada nas seguintes etapas: (a) caracterização das situações envolvendo argumentos 1 complexos; (b) especificação e implementação de padrões para convertê-los em argumentos simples, com a conversão da frase verbal em um frase substantiva; (c) caracterização das situações envolvendo argumentos 2 e trechos descartados complexos, para os quais foi necessário especificar regras de conversão para gerar novamente as extrações de relações a partir da sentença original, para evitar a perda da semântica da sentença original; (d) geração das regras de conversão a partir das situações identificadas; e (e) implementação de um conversor genérico, baseado em regras definidas em XML para realizar as extrações, de forma que os argumentos não utilizassem frases verbais. O corpus utilizado, para especificação, implementação e teste do protótipo, foi a seção “Food Choices” de um livro sobre nutrição básica, cujo texto expressa motivações sociais, religiosas e pessoais para escolha de alimentos, acarretando maior de dificuldade na extração de relações. Foram identificadas seis situações distintas com necessidade de conversão de argumentos 1. Foi especificada uma notação baseada preferencialmente em categorias de frases (nominais, verbais, adverbiais, etc) contidas nas sentenças para denotar regras genéricas de extração de relações binárias, para tratar argumentos 2 complexos e trechos descartados complexos. Como resultado deste trabalho, as extrações binárias geradas, conservam a semântica da sentença original. A automação da aplicação de regras de extração definidas em XML potencializa a prospecção futura de um conjunto de regras mais abrangente, para servir como ponto de partida para um processo automatizado de geração das regras de conversão, tornando o processo como um todo de extração de relações binárias totalmente automatizado.