

Witold Kosiński (P JWSTK)

Tomasz Kuśmierczyk (IPI PAN)

Marcin Sydow (P JWSTK, IPI PAN)

E-mail: wkos@poljap.edu.pl, msyd@ipipan.waw.pl

Podsumowania węzłów w semantycznych grafach wiedzy jako problem optymalizacyjny

Semantyczne grafy wiedzy są zdobywającym popularność formatem do reprezentowania i wyszukiwania danych semantycznych. Węzły w takich grafach reprezentują encje (byty) z jakiejś dziedziny (np. pisarze, książki, etc.), a skierowane krawędzie stanowią instancje binarnych relacji pomiędzy encjami, które reprezentują fakty (np. że dany autor napisał daną książkę, etc.).

W problemie podsumowania węzła chodzi o wybór niewielkiego zbioru krawędzi S , o ustalonym rozmiarze, incydentnych z danym węzłem, aby przedstawić jak najbardziej reprezentatywną informację o danym węźle. Problem przedstawiony będzie jako problem optymalizacyjny, gdzie maksymalizowana jest następująca funkcja celu:

$$obj(S) = (k - 1) \sum_{d \in S} imp(d) + 2\delta \sum_{d_1 \neq d_2 \in S} diss(d_1, d_2),$$

gdzie $imp : D \rightarrow Q^+$ jest miarą „wagi” krawędzi d , natomiast $diss : D^2 \rightarrow Q^+$ reprezentuje liczbowo „niepodobieństwo” pary krawędzi, a δ jest parametrem. Innymi słowy, chodzi o wybór „ważnych” faktów, które są parami niepodobne do siebie. Wyjaśnione będzie, że ww. problem optymalizacyjny jest NP-trudny.

Zaprezentowane też będą wstępne wyniki eksperymentalne z zastosowania algorytmu mrówkowego z nowym mechanizmem samoadaptacji do efektywnego rozwiązania ww. problemu.