Preferences, Links, and Probabilities for Ranking Objects in Ontologies

Thomas Lukasiewicz [1] [2]   Jörg Schellhase [3]

[1] Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza", Italy

[2] Institut für Informationssysteme
Technische Universität Wien, Austria

[3] Fachgebiet Wirtschaftsinformatik
Universität Kassel, Germany

Outline

- Preference Ranking
- Google's PageRank
- Combined Ranking
- Probabilistic Ontologies

Ranking data objects in web / product / literature search:

Given a query Q (strict and desired properties) and a set of data objects O, return a ranking on O relative to the query Q.

Conditional preference bases (L, P) as a means for ranking objects in ontologies (Lukasiewicz & Schellhase, ESWC-06):

- L is a finite set of description logic concepts (strict);
- P is a finite set of conditional preferences $(\alpha|\phi)[s]$ (desired with strength s): "generally, in the context ϕ, property α is preferred to property $\neg\alpha$ with strength s".

Example (Literature Search): We are looking for publications with the word "matching" in the title; and we prefer conference papers to non-conference papers, and papers of international conferences to papers of national conferences:

$$\begin{aligned} L &= \{\mathit{Publication}, \mathit{in\_title}(\text{“}\mathit{matching}\text{”})\}\,, \\ P &= \{(\mathit{type}(\text{“}\mathit{international}\text{”}) | \mathit{ConferencePublication})[70]\,, \\ &\quad\; (\mathit{ConferencePublication})[80]\}\,. \end{aligned}$$

Query Q divides the publications in the query result into three groups: first international conference publications (lowest rank), second national conference publications (second lowest rank), and third non-conference publications (highest rank).

An orthogonal way of ranking data objects is based on the analysis of the link structure between the objects:

The PageRank of a web page u is defined as

$$R(u) \;=\; c \cdot (\textstyle\sum_{v \in B_u} R(v) \,/\, N_v + E(u))\,,$$

where (i) B_u is the set of pages that point to u, (ii) N_v is the number of links from v, (iii) c is a normalization factor, and (iv) $E(u)$ is a vector over web pages representing a source of rank.

Intuitively, the more web pages with high rank point to a web page, the higher is the rank of this web page.

We propose an approach in which our preference ranking is orthogonally combined with PageRank's importance ranking:

- (A) preference ranking as input $E(u)$ to the PageRank computation: this allows for influencing the PageRank ranking by user-defined conditional preferences, e.g.,
  - for a better web search, or
  - for personalization purposes;
- (B) refining the ranking based on user-defined conditional preferences by PageRank's importance ranking.

In order to rank partial objects, we additionally exploit the information encoded in probabilistic ontologies:

For example, suppose "every publication is a conference publication with probability 0.9". Thus, if we know that an object o is a publication, then we can conclude that it is a conference publication with probability 0.9, which can then be exploited to compute the (expected) rank of o.

The ranking κ^{sum} is defined as follows for all total objects o:

$$\kappa^{sum}(o) = \begin{cases} \infty & \text{if } o \not\models L \\ \sum_{(\alpha|\phi)[s] \in P^\star :\ o \models \phi \sqcap \neg\alpha} s + 1 & \text{otherwise.} \end{cases}$$

The ranking κ^{sum} is defined as follows for all partial objects o:

$$\kappa^{sum}(o) = \begin{cases} \infty & \text{if } o \not\models L \\ \sum_{(\alpha|\phi)[s] \in P^\star} Pr(\phi \sqcap \neg\alpha \mid o) \cdot s + 1 & \text{otherwise.} \end{cases}$$