Exploitation of User Profile in Travel Support System – Ontology-based Filtering Algorithm

Maciej Gawinecki

August 23, 2006

1 Formal definition

Let us assume that r is the node in the RDF graph and st is the triple (statement) in this graph. We say that *r belongs to st* if and only if it is either its *subject*, *predicate* or *object*. Formally we define this:

$$r \in st = (s, p, o) \Leftrightarrow r = s \vee r = p \vee r = o \tag{1}$$

Let's suppose that R is such r node, which is of *resource* type (not literal). Then we define *R-subgraph* as a set of these statements, which can be visited in RDF graph, travelling along objects of statements and starting from R node. You can imagine, that R is the root node of Restuarant individuals and we are trying to find whole description of this Restaurant. Formally, we define it in inductive manner:

$$\Pi_R^0 := \{st_0 : st_0 = (R, p_0, o_0)\} \tag{2}$$

$$\Pi_R^{i+1} := \{st_{i+1} : st_{i+1} = (o_i, p_{i+1}, o_{i+1}) \wedge o_i \text{ is of resource type}\} \tag{3}$$

$$\Pi_R := \bigcup_{i \geq 0} \Pi_R^i \tag{4}$$

Then we can calculate how many times chosen node r occurs in R-subgraph:

$$stat(r) = \#\{r : r \in st \wedge st \in \Pi_R\} \tag{5}$$

However, we are interested in these nodes (concepts), which we state opinions about in a user profile. These opinions reference to the concepts described in Restaurant domain ontology, not in Restaurant database. Therefore we remind the idea of reflection. Reflection of concept r into domain ontology D (often called schema ontology or T-Box) will be such function, which assigns to the r one of the following values (marked as $\overline{r}$):

- r if $r \in D$,
- C, if $r \notin D$ and r is instance of $C \in D$,

- otherwise r is undefined.

Then we can calculate number of reflected forms of r occuring in R-subgraph:

$$stat(\overline{r}) = \#\{\overline{r} : r \in st \wedge st \in \Pi_R\} \tag{6}$$

As you will see soon this formal introduction will help us to define ”temperature” of potential interest in the particular Restaurant. We define the temperature of the Restaurant pointed by the root node R as total sum of temperatures of nodes contained in R-subgraph:

$$temp(R) := \sum_{\overline{r} \in \overline{T_R}} temp(\overline{r}) \tag{7}$$

where T_R:

$$T_R := \{r : r \in st \wedge st \in \Pi_R\} \tag{8}$$

and its reflected form $\overline{T_R}$:

$$\overline{T_R} := \{\overline{r} : r \in T_R\} \tag{9}$$

The temperature of a single reflected node $\overline{r}$ is function of ”temperature” of opinion about this node, the number of occurences of $\overline{r}$ in R-subgraph and the fact whether $\overline{r}$ belongs to the user context or not:

$$temp(\overline{r}) := \begin{cases} \beta \cdot stat(\overline{r}) \cdot temp(opinion_{\overline{r}}) & \text{for } \overline{r} \in context \\ stat(\overline{r}) \cdot temp(opinion_{\overline{r}}) & \text{for } \overline{r} \notin context \end{cases} \tag{10}$$

β is predifined constant such $\beta \geq 1$. For explanation, context is simply user query in form of list of contraints, i.e. pairs of property and it expected value. We say $\overline{r}$ belongs to context if $\overline{r}$ is equal to either one of the properties or one of the values.

The ”temparature” of opinion about particular concept $\overline{r}$ depends on how the concept has been classified during profile learning phase and if has not been classified, then normalized probablity ($p^n_{\overline{r}}$) for this concept matters:

$$temp(opinion_{\overline{r}}) := \begin{cases} 1 & \text{if } \overline{r} \text{ classified as } \textit{interesting} \\ 0 & \text{if } \overline{r} \text{ classified as } \textit{not interesting} \\ & \text{or no opinion for } \overline{r} \\ \alpha \cdot p^n_{\overline{r}} & \text{otherwise} \end{cases} \tag{11}$$

where α is predefined constant, such $\alpha \in (0,1)$. Please also note, that user is not expected to be interested in particular concept if she has no opinion about.

When all Restaurants has been rated by assigment of proper ”temperature”, we would recommend to the user only the hottest ones, which formally means, they should have temparature over particular level L:

$$Output := \{R : R \in Results \wedge temp(R) \geq L\} \tag{12}$$

We define L as *expected value* over whole population of rated restaurants:

$$L := \mathrm{E}X := \sum_{R_i \in Results} temp(R_i) \cdot p_i \tag{13}$$

where X is random variable for the population. We can simply calculate, that $p_i := \frac{1}{\#Results}$.

2 Algorithm procedure

1. For whole set of results rate each restaurant identified by root node$R$ accordingly to the definition of temperature given above.
    (a) Calculate statistics (number of occurence) for reflection of each resource in R-subgraph.
    (b) Compute temperature for each R-subgraph.
2. Compute level L for rated restaurants.
3. Throw away restaurants with temperature below the level L.
4. Sort restaurants descendingly according to the assigned temperatures.

3 Consideration

This article is correction of method proposed in (Gawinecki, 2005) and inspired by works of (Ganzha *et al.*, 2006; Gawinecki *et al.*, 2005).

References

GANZHA, MARIA, GAWINECKI, MACIEJ, PAPRZYCKI, MARCIN, GASIOROWSKI, RAFAL, PISAREK, SZYMON, & HYSKA, WAWRZYNIEC. 2006. *Semantic Web Technologies and eBusiness: Virtual Organization and Business Process Automation*. Idea Publishing Group. Chap. Utilizing Semantic Web and Software Agents in a Travel Support System.

GAWINECKI, MACIEJ. 2005 (December). *Modelling User on the Basis of Interactions with a WWW Based System*. M.Phil. thesis, Adam Mickiewicz University, Poznan, Poland.

GAWINECKI, MACIEJ, KRUSZYK, MATEUSZ, & PAPRZYCKI, MARCIN. 2005. Ontology-based Stereotyping in a Travel Support System. *In: Proceedings of the XXI Fall Meeting of Polish Information Processing Society*.