

A Semantic Matchmaking System for Online Dating

Emily Wilson / Kenneth B. Kent
University of New Brunswick
Faculty of Computer Science
l4re6@unb.ca ken@unb.ca

Outline

- Application of semantic web tools such as ontology languages and reasoning software to an existing online dating website model
- Integration of geographical datasets used to infer new properties about customers and their potential suitors providing a more accurate match
- A four level matching architecture

Motivation

People may seek out online dating websites as a means of finding potential partners more efficiently. Producing a set of matches for a person that they will deem suitable is one of the major challenges in the online dating industry. Users enter information about themselves that is used to identify other users that are considered appropriate matches. However, users may deem the list of potential matches that they are returned as unsuitable. This leads to customers cancelling their memberships and giving the company's website poor reviews.

In this contribution we present an application of semantic technologies to online dating matching techniques. Information provided is used not only to eliminate matches deemed unsuitable but also to infer additional properties about the user, that will be used in the matching process.

Objectives

- Model geographical information in an ontology and integrate aspects of that information into the matching process.
- Create rules to be used by a reasoner to expose new data by reasoning over user data and geographical data.
- Improve matching results based on details extracted from user input to provide matches of higher semantic quality.
- Apply this matchmaking system to an online dating website

Methodology

The matching system uses a level based architecture. There are four levels in total. Matches are generated for users in the system with a score based on level. Members answer questions from numerous levels (depending on their membership), and the system performs the matching process based on a pre-defined set of rules in the form of background knowledge.



Level 1 matches a user based on their basic information (gender, looking for, etc.). Any other users that do not meet the level 1 requirements for a match are eliminated from the potential matches. Next, a user is matched by their answers to the questions in level 2 regarding priorities and given a score representing the accuracy of the match. Information from level 2 is then used in level 3 and 4 matching.

Level 4 data is inferred based on a user's geographical area. From public geographic datasets we can infer things like:

- Their housing
- Their income level
- Their language
- Their level of education
- Their employment status and occupation

Implementation

All user information in the system is represented by RDF triples. Thus if a user is from Fredericton then there exists an RDF triple in the knowledge base about that user of the form:

```
userid    city    "Fredericton"
```

Information from geographical datasets is also transformed into RDF triples. For example, the RDF triple to represent the primary language used in Fredericton has the form:

```
"Fredericton" primary-language "English"
```

General purpose rules are written and used by a reasoner to link the two datasets. By linking the RDF statements above the following RDF triple can be inferred:

```
userid    primary-language    "English"
```

The inferred triples are used in a 4th level of matching. Each level uses custom SPARQL queries to find a user's set of matches which are then returned.

Commercial Partner:

ATWOOD
TECHNOLOGY

UNIVERSITY OF NEW BRUNSWICK

Reconfigurable Computing Research Group

FACULTY OF COMPUTER SCIENCE

