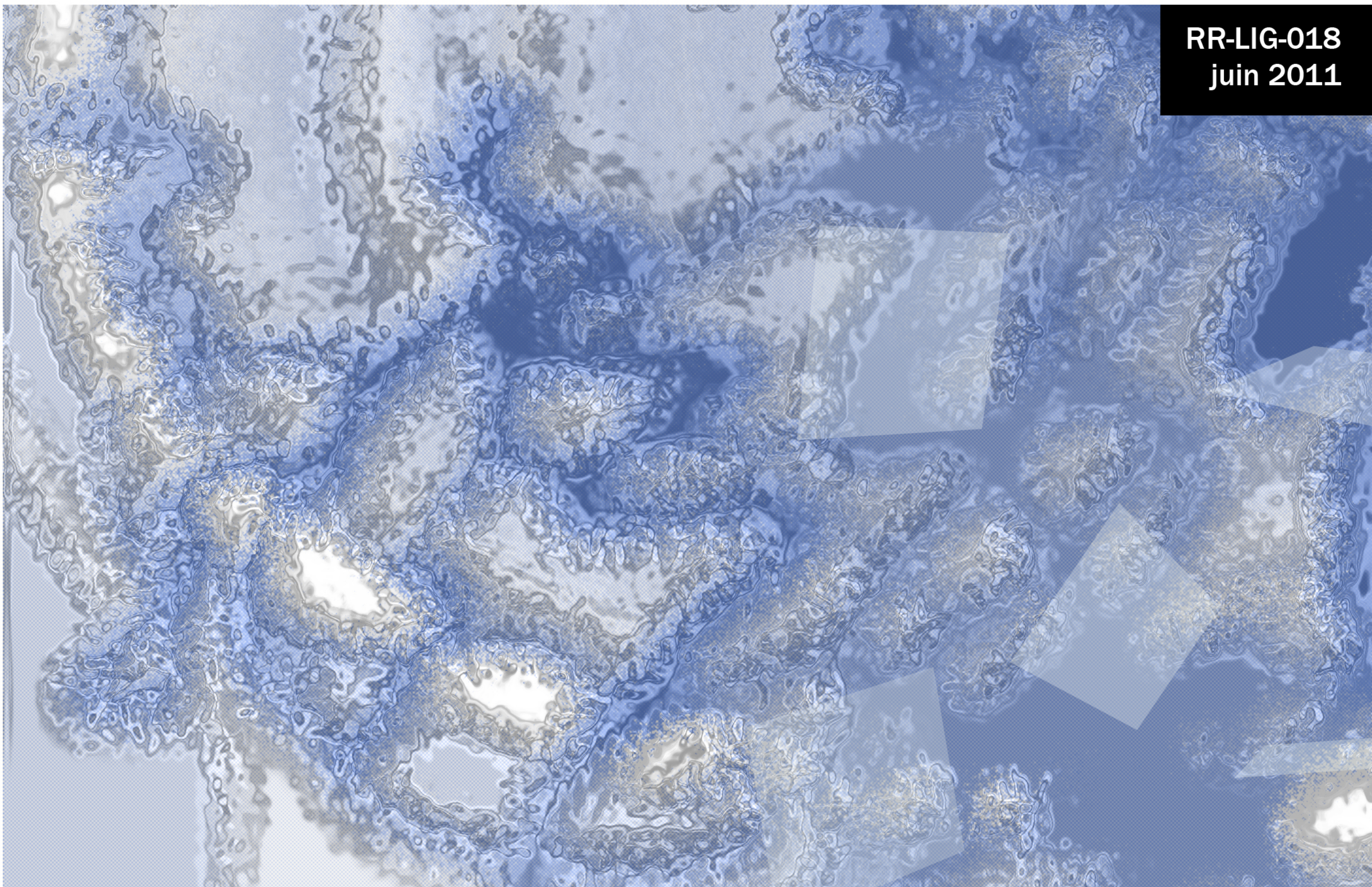
Laboratoire d'Informatique de Grenoble

Les rapports de recherche du LIG

Exploiting ontologies and alignments for trust in semantic P2P networks

Manuel ATENCIA, Post doc, LIG, Grenoble University (UJF), France
Jérôme EUZENAT, Senior Research Scientist, LIG, INRIA, France
Marie-Christine ROUSSET, Professor, LIG, Grenoble University (UJF), France



RR-LIG-018
juin 2011

http://rr.liglab.fr
ISSN 2105-0422

LIG

Exploiting ontologies and alignments for trust in semantic P2P networks

Manuel Atencia
Université de Grenoble 1 & INRIA

Jérôme Euzenat
INRIA

Marie-Christine Rousset
Université de Grenoble 1

Laboratoire d'informatique de Grenoble (LIG), Grenoble, France
Manuel.Atencia@inria.fr Jerome.Euzenat@inria.fr Marie-Christine.Rousset@imag.fr

Abstract

In a semantic P2P network, peers use separate ontologies and rely on alignments between their ontologies for translating queries. However, alignments may be limited —unsound or incomplete— and generate flawed translations, and thereby produce unsatisfactory answers. In this paper we propose a trust mechanism that can assist peers to select those in the network that are better suited to answer their queries. The trust that a peer has towards another peer is subject to a specific query and approximates the probability that the latter peer will provide a satisfactory answer. In order to compute trust, we exploit the information provided by peers' ontologies and alignments, along with the information that comes from peers' experience. Trust values are refined over time as more queries are sent and answers received, and we prove that these approximations converge.

Introduction

Recently, P2P systems have received considerable attention because their underlying infrastructure is appropriate to scalable and flexible distributed applications over Internet. In P2P systems, there is no centralized control or hierarchical organization: each peer is equivalent in functionality and cooperates with other peers in order to solve a collective task. P2P systems have evolved from simple keyword-based file sharing systems like Napster and Gnutella to semantic data management systems like EDUTELLA (Nejdl et al. 2002), PIAZZA (Halevy et al. 2003), or SOMEWHERE (Adjiman et al. 2006).

In this paper, by *semantic P2P networks* we refer to fully decentralized overlay networks of people or machines (called peers) sharing and searching various resources (documents, videos, photos, data, services) based on their semantic annotations using ontologies. In semantic P2P systems, each peer freely organizes its local resources as instances of classes of its own ontology serving as query interface for other peers. Alignments between ontologies make possible to reformulate queries from one local peer vocabulary to another. The result of a query is a set of resources (e.g., documents) that are instances of some classes corresponding, possibly through subsumption or alignment, to the initial query posed to a specific peer.

Trust is widely acknowledged as an important factor when considering networks of autonomous interacting entities and notably in the context of the Semantic Web. When referring to the notion of trust, T. Berners-Lee advocates for a user to be able to check for reasons why she could be confident with a returned answer (Berners-Lee 1997). Trust is helpful to select, from a given set of peers, the ones that will answer with most satisfactory instances. Peers may use this information for broadcasting their queries to a reduced set of peers and to have an approximation of the reliability of provided answers. In addition, peers may preventively send selected queries in order to improve the trust they have in another peer. Finally, by identifying "weak correspondences", peers may signal faulty alignments and trigger new matching of the ontologies.

Several proposals have already been made that do not all share the same meaning for trust (Sabater and Sierra 2005; Artz and Gil 2007). Many of them are user/agent/peer centered and rely on the assumption that all peers share similar implicit goals. Trust is then closely related to the notion of reputation in a community.

In the context of semantic P2P systems, peers may correspond to different points of view. For this reason, we rather promote the computation of subjective trust values based on direct experiences between peers. We also argue for a finer grained (or context sensitive) approach to trust in order to take into account the fact that, for answers provided by a same peer, the trust into these answers may vary according to which class they are instance of within the peer ontology.

We propose a probabilistic model to handle trust in a P2P setting. The trust of a peer P towards another peer P' regarding a class C (belonging to the ontology of P) approximates the probability that an instance returned by P' (possibly through a path of alignments) as an answer to the query asking for instances of C is satisfactory for P. In order to compute trust, we exploit the information provided by peers' ontologies and alignments, along with the information that comes from direct experience. Trust values are refined over time as more queries are sent and received, and we prove that these approximations converge.

The paper is organized as follows. We first present the background of our work. Then we introduce our probabilistic definition of trust of a peer into another regarding a query. We show how it is possible to estimate it by means of the Bayesian approach to statistics, which consists in taking into account feedbacks on past experiences. Finally we compare our approach with related work on handling trust in P2P or ad-hoc networks, and give some perspectives.

Preliminaries

In this section we define the components of a semantic P2P network: ontologies, alignments and acquaintance graphs. The kind of queries that peers will pose is also described.

Ontologies and Populated Ontologies

We draw a distinction between the ontological structure and the instances used to populate it. We deal with lightweight ontologies: classes linked by means of a less-general-than relationship and a disjointness relationship.

Definition 1 *An* ontology *is a tuple* $O = \langle C, \leqslant, \perp \rangle$ *where* C *is a non-empty finite set of* classes*;* $\leqslant$ *is a partial order on* C*;* $\perp$ *is an irreflexive and symmetric relation on* C*; and for all* $c, c', d, d' \in C$,

$$\text{if } c \perp d, c' \leqslant c \text{ and } d' \leqslant d \text{ then } c' \perp d'.$$

A populated ontology is the result of adding instances to an ontology in accordance to the intended meaning of the two ontological relationships.

Definition 2 *A* populated ontology $\mathcal{O}$ *is a tuple* $\langle O, I, ext \rangle$*, where* O *is an ontology,* I *is a set of* instances*, and* ext *is a function that maps each class* c *of* O *with a subset* $ext(c)$ *of* I *called the* extension *of* c *in such a way that the family of class extensions covers* I*, and for all classes* c, d *the following properties hold:*

1. *if* $c \leqslant d$ *then* $ext(c) \subseteq ext(d)$*, and*
2. *if* $c \perp d$ *then* $ext(c) \cap ext(d) = \emptyset$*.*

Alignments

In an open and dynamic environment as a P2P network, the assumption of all peers sharing the same ontology is not realistic. But if peers fall back on different ontologies, there must be a way to connect ontologies and translate queries so that their addressees are able to process them. Typically this is done by means of alignments —sets of correspondences between semantically related ontological entities— and the operation of finding alignments is called ontology matching (Euzenat and Shvaiko 2007).

A correspondence between two classes c and c' of two ontologies O and O', respectively, is usually defined as a tuple $\langle c, c', r \rangle$ with $r \in \{\leqslant, =, \geqslant\}$, and $c \leqslant d$ (or $\langle c, c', \leqslant \rangle$) is read "$c$ is less general than d", $c \geqslant d$ is read "c is more general than d", and $c = d$ is read "c is equal or equivalent to d". Here, however, we deal with a more general notion of a correspondence inspired from (Euzenat 2008).

Definition 3 *Let* O *and* O' *be two ontologies, and let* c *and* c' *be classes of* O *and* O'*, respectively. A* correspondence *between* c *and* c' *is a tuple* $\langle c, c', R \rangle$ *with* $R \in 2^{\Gamma} \setminus \emptyset$ *where* Γ *is the set* $\{=, >, <, \between, \perp\}$*. An* alignment $\mathcal{A}$ *between* O *and* O' *is a set of correspondences between classes of* O *and* O'*.*

The above definition requires a clarification. Firstly, a class is not connected to another one via a relation, but a set of relations, and this set is to be thought of as an exclusive disjunction. For instance, $c\{>, <\}d$ (that is, $\langle c, d, \{>, <\} \rangle$) is read "either c is more general than d or less general than d". In this way, we can express uncertainty with regard to the alignment relation. Notice that '$\geqslant$' and '$\leqslant$' can be seen as abbreviations for $\{=, >\}$ and $\{=, <\}$, respectively. Also a non-standard symbol '$\between$' is introduced. It reflects the idea of overlapping: classes the extensions of which share some instances but no one is equal to or contained into the other. Finally, $c\,\Gamma\, d$ states total uncertainty on the relation between the classes c and d.

There exist various ways to compute alignments, e.g., manual or automatic matching, composing or inverting other alignments. This is out of the scope of this paper and readers are referred to (Euzenat and Shvaiko 2007; Euzenat 2008).

Acquaintance Graphs

We consider a set $\mathcal{P} = \{P_i\}_{i=1}^n$ of peers. In this paper peer P_i is identified by i. We assume that peer P_i is associated with one populated ontology denoted by $\mathcal{O}_i = \langle O_i, I_i, ext_i \rangle$ (where $1 \leq i \leq n$).

An acquaintance graph stands for peers' acquaintances (or neighbors) in the network. An edge between two peers reflects the fact that they know the existence of each other and that there exists at least one alignment between their respective ontologies.

Definition 4 *An* acquaintance graph *is a labelled directed graph* $\langle \mathcal{P}, \text{ACQ} \rangle$ *where* $\mathcal{P} = \{P_i\}_{i=1}^n$ *is the set of vertices and any edge in* ACQ *is of the form* $\langle i, j \rangle$ *with* $i \neq j$*, and it is labelled with a non-empty set* $\mathcal{A}(i, j)$ *of alignments between* O_i *and* O_j*. Peer* P_j *is said to be an* acquaintance *of peer* P_i *provided that* $\langle i, j \rangle \in$ ACQ*.*

We do not impose $\mathcal{A}(i, j)$ to be singleton as peers may resort to different matchers or compute alignments by both composition and inverse operations. In a semantic P2P network, peers' links continually appear and disappear: acquaintance graphs are under constant change.

If two peers P_i and P_j know the existence of each other and they share an ontology $O_i = O_j$, $\langle i, j \rangle$ can be labelled with the identity alignment (the one that is made up of all correspondences $\langle c, c, = \rangle$ with $c \in C_i = C_j$). Note, though, that the populated ontologies $\mathcal{O}_i$ and $\mathcal{O}_j$ may be different.

Queries and Query Translations

Peers pose queries to obtain information concerning others' ontologies. In this work we deal with a simple query language, as peers can only request class instances: if peer P_j is an acquaintance of peer P_i, it may be asked

$$\mathcal{Q} = c(X)? \qquad (1)$$

by P_i with $c \in O_i$. Queries may require to be translated for their recipients to be able to process them. Query translations are determined by correspondences of the existing alignments in the network. Specifically, if P_i wants to send $\mathcal{Q}$ to P_j, it will firstly choose a correspondence $\langle c, d, R \rangle$ (if there exists one) of an alignment $\mathcal{A} \in \mathcal{A}(i,j)$ and then send P_j the translation

$$\mathcal{Q}' = d(X)? \tag{2}$$

In turn, P_j can forward $\mathcal{Q}'$ or send translations of it to some of its acquaintances, and so forth.

The answer to (1) through its translation (2), is the set of instances of d in P_j's populated ontology. Unlike queries, we assume that no translation of instances is ever required. Since alignments may be unsound or incomplete, this answer may contain *unsatisfactory* instances, i.e., instances which are not considered instances of c by P_i (even if R is the equality relation).

A peer cannot foresee whether the answer that another provides to one of its queries contains satisfactory instances or not, but this uncertainty can be estimated with the help of a trust mechanism.

The Trust Mechanism

We look at trust as a means to estimate the proportion of satisfactory instances in an answer. The idea of satisfactory instance can be faithfully captured by a populated ontology $\mathcal{O}_i^*$ that gives account of a hypothetical situation in which peer P_i had classified all instances of the network respecting its ontology O_i. In this way we can express the fact that P_i considers an arbitrary instance a as an instance of $c \in O_i$ by means of the more succinct choice $a \in ext_i^*(c)$. It is assumed that $O_i = O_i^*$ and $ext_i(c) \subseteq ext_i^*(c)$ for every class $c \in C_i$.

If peer P_i receives a set B of instances as an answer to the query (1), the proportion of satisfactory instances is given by the conditional probability $p(ext_i^*(c)|B)$. The probability space under consideration is the triple $(\Omega, \mathfrak{A}, p(\cdot))$ where Ω is the set of instances of the network (a finite set), $\mathfrak{A}$ is the power set of Ω, and $p(\cdot)$ is Laplace's definition of probability. Our approach for trust aims at approximating these conditional probabilities. Trust values can be used by peers to select from a set of potential addresses those which are expected to answer a query with most satisfactory instances.

Before the definition of trust, we introduce the notion of a probabilistic populated ontology.

Probabilistic Populated Ontologies

Once an answer is received, it is added to the extension of the queried class. In order to capture the evolvement of class extensions in the network, we consider a discrete time variable $t \in \mathbb{N}$, and we write $\mathcal{O}_i^t$ to denote peer P_i's populated ontology at instant t, beginning with $\mathcal{O}_i = \mathcal{O}_i^0$:

$$\mathcal{O}_i = \mathcal{O}_i^0, \mathcal{O}_i^1, \ldots, \mathcal{O}_i^t, \ldots \tag{3}$$

It is assumed that the underlying ontology does not change ever, that is, $O_i = O_i^t$ for all $t \in \mathbb{N}$, and that $\{ext_i^t(c)\}_{t \in \mathbb{N}}$ is a non-decreasing sequence for all $c \in C_i$.

However, since we deal with approximations of probabilities, new instances may not be 100% satisfactory. For this reason, at time t, P_i is associated with a *probabilistic populated ontology*.

Peer P_i's probabilistic populated ontology at time t is a triple

$$\widetilde{\mathcal{O}}_i^t = \langle O_i, I_i^t, \widetilde{ext}_i^t \rangle$$

such that I_i^t is a set of instances and $\widetilde{ext}_i^t$ is a function that maps each class c of O_i with its *probabilistic extension*

$$\widetilde{ext}_i^t(c) = \langle A^*, \Delta \rangle$$

where

- A^* is a (possibly empty) subset of $ext_i^*(c)$, i.e., a set of instances that are certainly instances of the class c, and
- Δ is empty or a collection $\{\langle A^k, U^k \rangle\}_{k=1}^m$ where all A^k are pairwise disjoint subsets of I_i^t which are also disjoint from A^*, and all U^k are different real intervals of the form $[p^k, p^k]$ or $[p^k, 1]$, where $p_k \in (0,1]$ is an approximation or an approximated lower bound of the probability for A^k to be a set of instances of the class c:

$$p(ext_i^*(c)|A^k) \approx p_k \qquad \text{if } U^k = [p^k, p^k]$$
$$p_k \lesssim p(ext_i^*(c)|A^k) \leq 1 \qquad \text{if } U^k = [p^k, 1]$$

Further, if we define the extension $ext_i^t(c) = A^* \uplus \biguplus_{k=1}^m A^k$, the tuple $\mathcal{O}_i^t = \langle O_i, I_i^t, ext_i^t \rangle$ must be a (classical) populated ontology (in the sense that the axioms that relate classes with their extensions are fulfilled).

The use of real intervals instead of real numbers follows Lukasiewicz's notation for conditional constraints in probabilistic knowledge bases (Lukasiewicz and Straccia 2008).

Remark 1 *Every peer populated ontology $\mathcal{O}_i$ can be seen as a probabilistic populated ontology $\widetilde{\mathcal{O}}_i = \langle O_i, I_i, \widetilde{ext}_i \rangle$ where $\widetilde{ext}_i(c) = \{\langle ext_i(c), \emptyset \rangle\}$ for all $c \in C$.*

Peers build probabilistic populated ontologies as more queries are sent and answered, starting with the "probabilistic version" of $\mathcal{O}_i$:

$$\widetilde{\mathcal{O}}_i = \widetilde{\mathcal{O}}_i^0, \widetilde{\mathcal{O}}_i^1, \ldots, \widetilde{\mathcal{O}}_i^t, \ldots \tag{4}$$

And what was said about (3) at the beginning of this section holds for the underlying populated ontologies of (4).

Definition of Trust

With the new terminology, peer P_j's answer to the query (1) through its translation (2) is the probabilistic extension

$$\widetilde{B} = \widetilde{ext}_j^t(d)$$

and if $ext_j^t(d)$ is denoted by B, an arbitrary instance $a \in B$ is qualified as a satisfactory instance providing $a \in ext_i^*(c)$. The proportion of satisfactory instances in B is given by the conditional probability $p(ext_i^*(c)|B)$. Peer P_i trusts peer P_j as much as this value is high.

Definition 5 *Let us consider two peers P_i and P_j ($i \neq j$) and let $\langle c, d, R \rangle$ be a correspondence of an alignment of $\mathcal{A}(i,j)$. The* trust *that peer P_i has towards P_j at time t with regard to the correspondence $\langle c, d, R \rangle$ will be denoted by*

$$trust^t(P_i, P_j, \langle c, d, R \rangle)$$

and approximates the probability

$$p(ext_i^*(c)|ext_j^*(d), ext_j^t(d))$$

Notice that $ext^*_j(d)$ has been included in the "given" part. In this way, P_i can benefit from peer P_j's confidence on its own instances. This leads us to the notion of self-confidence.

Definition 6 *The* self-confidence *of peer P_i with regard to a class $c \in C_i$ at time t will be denoted by $self^t(P_i, c)$ and approximates the probability $p(ext^*_i(c)|ext^t_i(c))$.*

Both definitions are incomplete as we do not specify how the probabilities are approximated. This comes below where we explain the computation of trust.

Computation of Trust

Without loss of generality, we assume that no class extension in any peer ontology $\mathcal{O}_i$ is empty. The following theorem is the basis for completing Definition 6.

Theorem 1 *Assume that $\widetilde{ext}^t_i(c) = \langle A^*, \{\langle A^k, U^k\rangle\}^m_{k=1}\rangle$. Let $N = |ext^t_i(c)|$, $N^* = |A^*|$, $N^k = |A^k|$ for each k, and*

$$s = \frac{1}{N}\big(N^* + \sum_{k=1}^{m} p_k N^k\big)$$

If there exists at least a k_0 with $U^{k_0} = [p^{k_0}, 1]$ and $p^{k_0} \neq 1$, $s \lesssim p(ext^*_i(c)|ext^t_i(c)) \leq 1$, if not, $p(ext^*_i(c)|ext^t_i(c)) \approx s$.

Proof: it all boils down to the fact that $p(ext^*_i(c)|A^*) = 1$, and hence we have $p(ext^*_i(c)|ext^t_i(c)) = p(A^*|ext^t_i(c)) + \sum_{k=1}^m p(ext^*_i(c)|A^k)p(A^k|ext^t_i(c))$.

Peer P_i's self-confidence is defined as s or the midpoint of the interval $[s, 1]$:

$$self^t(P_i, c) =_{def} \begin{cases} \frac{1}{2}(1+s) & \text{if there exists a such } k_0 \\ s & \text{otherwise} \end{cases}$$

Notice that, as one would expect, $self^0(P_i, c) = 1$ for all c of P_i's ontology.

Regarding Definition 5, two kinds of information will be considered when computing trust: the information provided by alignments —gathered in a single trust value referred to as *alignment-based* trust and denoted by $trust_a$— and peers' direct experience —or *direct* trust denoted by $trust^t_d$. Trust is then defined as a convex combination of these two values ($\langle P_i, P_j, \langle c, d, R\rangle\rangle$ is omitted for the sake of simplicity):

$$trust^t =_{def} \lambda^t_a \cdot trust_a + \lambda^t_d \cdot trust^t_d \quad (5)$$

where λ^t_a and λ^t_d are the *reliabilities* on $trust_a$ and $trust^t_d$ at time t, respectively. In what follows we introduce all these elements. Before, we present some helpful results from probability theory the proofs of which are straightforward.

Lemma 1 *Let $(\Omega, \mathfrak{A}, p)$ be a probability space. Let A, B and C be three events with $p(A, B) > 0$. Then*

$$p(C|A)+p(B|A)-1 \leq p(C|A, B) \leq p(C|A)-p(B|A)+1$$

Lemma 2 *Let $(\Omega, \mathfrak{A}, p)$ be a probability space. Let A, B and C be three events with $p(A \setminus B) > 0$. Then*

$$p(C|A \setminus B) \geq p(C|A) - p(B|A)$$

Alignment-based trust. If there is no past experience at all, neither from P_i nor from its aquaintances, P_i can only rely on the information provided by the alignments. If these were correct then

$$\begin{array}{lll} c\{=\}d & \text{iff} & ext^*_i(c) = ext^*_j(d) \\ c\{>\}d & \text{iff} & ext^*_i(c) \supset ext^*_j(d) \\ c\{<\}d & \text{iff} & ext^*_i(c) \subset ext^*_j(d) \\ c\{\between\}d & \text{iff} & ext^*_i(c) \cap ext^*_j(d) \neq \emptyset \\ c\{\perp\}d & \text{iff} & ext^*_i(c) \cap ext^*_j(d) = \emptyset \end{array}$$

Now, recall that

$$p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) = \frac{p(ext^*_i(c), ext^*_j(d), ext^t_j(d))}{p(ext^*_j(d), ext^t_j(d))}$$

Therefore

$$\begin{array}{l} \text{if } c\{=\}d \text{ or } c\{>\}d \text{ then } p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) = 1 \\ \text{if } c\{<\}d \text{ or } c\{\between\}d \text{ then } p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) \in [0, 1] \\ \text{if } c\{\perp\}d \text{ then } p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) = 0 \end{array}$$

Our proposal is:

$$trust_a(P_i, P_j, \langle c, d, R\rangle) =_{def} \frac{1}{|R|}\sum_{r\in R} t(r)$$

such that:

$$t(r) = \begin{cases} 1 & \text{if } r \text{ is '='} \text{ or '>'} \\ \frac{1}{2} & \text{if } r \text{ is '<'} \text{ or '}\between\text{'} \\ 0 & \text{if } r \text{ is '}\perp\text{'} \end{cases}$$

Notice that if r is '$<$' or '$\between$' then $t(r) = \frac{1}{2}$, which is the mean of a uniform distribution in the interval $[0, 1]$. Also, since there is no information at all, we consider all relations in R to be equiprobable.

Since alignment trust is constant over time, we write $trust_a$ instead of $trust^t_a$ in (5).

Direct trust. The more queries are sent and received, the more the alignment-based trust can be contrasted. Direct trust relies on peers' direct experience so as to estimate the proportion of satisfactory instances in an answer. Moreover, it exploits the information included in peers' probabilistic populated ontologies. Let us explain how this is done.

First of all, by Lemma 1,

$$p^t_1 \leq p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) \leq p^t_2$$

where $p^t_1 = p(ext^*_i(c)|ext^t_j(d)) + p(ext^*_j(d)|ext^t_j(d)) - 1$, $p^t_2 = p(ext^*_i(c)|ext^t_j(d)) - p(ext^*_j(d)|ext^t_j(d)) + 1$. And with the help of Theorem 1, we can find an approximated lower bound s^t for $p(ext^*_j(d)|ext^t_j(d))$. Our goal is to find two real numbers such that

$$q^t_1 \lesssim p(ext^*_i(c)|ext^t_j(d)) \lesssim q^t_2$$

In this way, $\widetilde{p}^t_1 \lesssim p(ext^*_i(c)|ext^*_j(d), ext^t_j(d)) \lesssim \widetilde{p}^t_2$, where $\widetilde{p}^t_1 = q^t_1 + s^t - 1$ and $\widetilde{p}^t_2 = q^t_2 - s^t + 1$.

We define $trust_d^t$ as the midpoint of the interval $[\widetilde{p}_1^t, \widetilde{p}_2^t]$:

$$trust_d^t(P_i, P_j, \langle c, d, R\rangle) =_{def} \frac{\widetilde{p}_1^t + \widetilde{p}_2^t}{2}$$

In order to compute q_1^t and q_2^t we proceed by induction on t. For the lack of space we omit the basis and go straight to the inductive step: q_1^t and q_2^t are given and q_1^{t+1} and q_2^{t+1} are to be computed. The following lemma, which is a direct consequence of Theorem 1 and Lemma 1, will be helpful in this regard.

Lemma 3 *Let B be an arbitrary set of instances. Then*

$$p \lesssim p(ext_i^*(c)|ext_i^t(c), B) \lesssim q$$

where:

$$p = s + \frac{|B \cap ext_i^t(c)|}{|ext_i^t(c)|} - 1 \quad q = s - \frac{|B \cap ext_i^t(c)|}{|ext_i^t(c)|} + 1$$

and s is given by Theorem 1.

Without loss of generality we assume there is $B \neq \emptyset$ with $ext_j^{t+1}(d) = ext_j^t(d) \uplus B$. Thus $p(ext_i^*(c)|ext_j^{t+1}(d)) =$

$$p(ext_i^*(c)|ext_j^t(d))p(ext_j^t(d)|ext_j^{t+1}(d)) + p(ext_i^*(c)|B)p(B|ext_j^{t+1}(d))$$

Now, the set B can be partitioned into three subsets:

$$\begin{aligned}
\Im_{aut}^+ &= \{a \in B : a \in ext_i^t(c)\} = ext_i^t(c) \cap B \\
\Im_{aut}^- &= \{a \in B : \text{there exists } c' \in O_i \\
&\qquad \text{with } a \in ext_i^t(c') \text{ and } c \perp d\} \\
\Im_{\overline{aut}} &= B \setminus (\Im_{aut}^+ \uplus \Im_{aut}^-)
\end{aligned}$$

and since $B = \Im_{aut}^+ \uplus \Im_{aut}^- \uplus \Im_{\overline{aut}}$, then $p(ext_i^*(c)|B) =$

$$p(ext_i^*(c)|\Im_{aut}^+)p(\Im_{aut}^+|B) + p(ext_i^*(c)|\Im_{aut}^-)p(\Im_{aut}^-|B) + p(ext_i^*(c)|\Im_{\overline{aut}})p(\Im_{\overline{aut}}|B)$$

The sets $\Im_{aut}^+$ and $\Im_{aut}^-$ contain the instances of B that can be processed automatically: those that already belong to the extension of the query and the ones that are in extensions of classes that are disjoint from the query. In order to estimate $p(ext_i^*(c)|\Im_{aut}^+)$ and $p(ext_i^*(c)|\Im_{aut}^-)$ we resort to Lemma 3. There exists two numbers p_{aut}^+ and q_{aut}^+ such that

$$p_{aut}^+ \lesssim p(ext_i^*(c)|\Im_{aut}^+) \lesssim q_{aut}^+$$

Regarding $p(ext_i^*(c)|\Im_{aut}^-)$, we partition $\Im_{aut}^-$ into subsets $\{D^k\}_{k=1}^r$ such that for each k, $D^k \subseteq ext_i^t(c^k)$, where c^k is a maximal superclass that is disjoint from c. Therefore we have $ext_i^*(c^k) \cap ext_i^*(c) = \emptyset$ for all k and

$$\begin{aligned}
p(ext_i^*(c)|\Im_{aut}^-) &= 1 - p(\overline{ext_i^*(c)}|\Im_{aut}^-) \\
&\leq 1 - p(\bigcup_{k=1}^r ext_i^*(c^k)|\Im_{aut}^-) \\
&= 1 - \sum_{l=1}^r p(\bigcup_{k=1}^r ext_i^*(c^k)|D^l)p(D^l|\Im_{aut}^-) \\
&\leq 1 - \sum_{k=1}^r p(ext_i^*(c^k)|D^k)p(D^k|\Im_{aut}^-)
\end{aligned}$$

By Lemma 3 we can find a real numbers $\{p^k\}_{k=1}^r$ such that

$$p^k \lesssim p(ext_i^*(c^k)|D^k)$$

for each $k \in [1..r]$. Then $p(ext_i^*(c)|\Im_{aut}^-) \lesssim q_{aut}^-$ where

$$q_{aut}^- = 1 - \frac{1}{|\Im_{aut}^-|}\sum_{k=1}^r p^k|D^k|$$

The set $\Im_{\overline{aut}}$ contains instances that cannot automatically processed. To estimate the probability $p = p(ext_i^*(c)|\Im_{\overline{aut}})$ peer P_i is assumed to perform a sampling with replacement over the set $\Im_{\overline{aut}}$ and to decide the number of satisfactory instances with the help of an oracle. If $g > 0$ is the size of the sample G and h ($0 \leq h \leq g$) is the number of satisfactory instances then

$$p(ext_i^*(c)|\Im_{\overline{aut}}) \approx p_{\overline{aut}} = \frac{h+1}{g+2}$$

This value is the first moment of a beta distribution, as it is often used in Bayesian inference to describe the unknown parameter of a binomial distribution.

Finally, $q_1^{new} \lesssim p(ext_i^*(c)|B) \lesssim q_2^{new}$ with

$$\begin{aligned}
q_1^{new} &= \frac{1}{|B|}(p_{aut}^+ \cdot n_{aut}^+ + p_{\overline{aut}} \cdot n_{\overline{aut}}) \\
q_2^{new} &= \frac{1}{|B|}(q_{aut}^+ \cdot n_{aut}^+ + q_{aut}^- \cdot n_{aut}^- + p_{\overline{aut}} \cdot n_{\overline{aut}})
\end{aligned}$$

where $n_{aut}^+ = |\Im_{aut}^+|$, $n_{aut}^- = |\Im_{aut}^-|$ and $n_{\overline{aut}} = |\Im_{\overline{aut}}|$. And then $q_1^{t+1} \lesssim p(ext_i^*(c)|ext_j^{t+1}(d)) \lesssim q_2^{t+1}$ where

$$\begin{aligned}
q_1^{t+1} &= \frac{1}{|ext_j^{t+1}(d)|}(q_1^t \cdot |ext_j^t(d)| + q_1^{new} \cdot |B|) \\
q_2^{t+1} &= \frac{1}{|ext_j^{t+1}(d)|}(q_2^t \cdot |ext_j^t(d)| + q_2^{new} \cdot |B|)
\end{aligned}$$

Reliabilities of Trust As already hinted before, as time goes, we want direct trust to carry more weight in the global trust value than the alignment-based trust. In other words, $trust_a$ becomes less *reliable* than $trust_d^t$ as t gets higher. This is faithfully captured by two non-negative functions λ_a^t and λ_d^t such that $\lambda_a^t + \lambda_d^t = 1$, and with $\lim_{t\to\infty} \lambda_a^t = 0$ and $\lim_{t\to\infty} \lambda_d^t = 1$. As an example:

$$\lambda_a^t = \frac{1}{t+1} \qquad \lambda_d^t = \frac{t}{t+1}$$

In this particular case, $\lambda_a^0 = 1$ and $\lambda_d^0 = 0$.

Updating Probabilistic Populated Ontologies

Notice that $\Im_{\overline{aut}}$ gathers together all new instances in P_j's answer. If G^+ and G^- are the subsets of satisfactory and unsatisfactory answers in the sample G then $B' = \Im_{\overline{aut}} \setminus G^-$ is to be part of the new extension of c. It is straightforward to see that $p(ext_i^*(c)|B') \approx q'$ where

$$q' = \frac{1}{|B'|}\big(p_{\overline{aut}} \cdot |\Im_{\overline{aut}} \setminus G| + |G^+|\big)$$

If $\widetilde{ext}_i^t(c) = \langle A^*, \{\langle A^k, U^k\rangle\}_{k=1}^m\rangle$, $\widetilde{ext}_i^{t+1}(c) = \langle B^*, \Theta\rangle$ where $B^* = A^* \uplus S^+$ and $\Theta = \Delta \uplus \langle B', [q', q']\rangle$ unless there exists k with $U^k = [q', q']$, in which case the pair $\langle A^k, U^k\rangle$ is just replaced with $\langle A^k \uplus B', [q', q']\rangle$.

In order for $\widetilde{\mathcal{O}}_i^{t+1}$ to be a probabilistic populated ontology, B' must be also included in the extension of any superclass of c. Let d be a superclass of c. Recall that $\mathfrak{I}_{aut}^- \cap \mathfrak{I}_{\overline{aut}} = \emptyset$, so B' does not contain any instance which belongs to the extension of a class disjoint from c. Hence it does not contain any instance that belongs to the extension of a class disjoint from d either. Actually it is not B', but $B'' = B' \setminus ext_i^t(d)$ the set to be added, so what we obtain is indeed a partition.

By Lemma 2,

$$\begin{aligned} p(ext_i^*(d)|B'') &\geq p(ext_i^*(d)|B') - p(ext_i^t(d)|B') \\ &\geq p(ext_i^*(c)|B') - p(ext_i^t(d)|B') \end{aligned}$$

Therefore $p(ext_i^*(d)|B'') \gtrsim q''$ where

$$q'' = q' - \frac{|B' \cap ext_i^t(d)|}{|B'|}$$

The new probabilistic extension $\widetilde{ext}_i^{t+1}(d)$ is built by adding $\langle B'', [q'', 1]\rangle$ in a similar way as before. It is easy to check that $\widetilde{\mathcal{O}}_i^{t+1}$ is a probabilistic populated ontology.

Convergence of Trust

In order to prove the convergence of $trust^t$, we only need to check $trust_d^t$, as $trust_a$ is constant and λ_a^t and λ_d^t converge to 0 and 1, respectively, as t approaches ∞.

In a previous section it is described the computation of $trust_d^{t+1}$ on the basis of $trust_d^t$ and an approximation of the probability of the new part B to be satisfactory —the interval $[q_1^{new}, q_2^{new}]$. This has a counter part s^{new} regarding P_j's self-confidences at time $t+1$ and t. The following is direct.

Lemma 4 *Suppose that $s^t \leq s^{new}$. Then*

1. *if $q_1^t \leq q_1^{new}$ then $\widetilde{p}_1^t \leq \widetilde{p}_1^{t+1}$, and*
2. *if $q_2^{new} \leq q_2^t$ then $\widetilde{p}_2^{t+1} \leq \widetilde{p}_2^t$.*

Theorem 2 *Assume that any time t we have $s^t \leq s^{new}$, and $q_1^t \leq q_1^{new}$ and $q_2^{new} \leq q_2^t$. Then $\{trust_d^t\}_{t\in\mathbb{N}}$ converges.*

Proof: under these hypotheses the collection $\{[p_1^t, p_2^t]\}_{t\in\mathbb{N}}$ is a system of nested intervals, and by Cantor's theorem

$$\bigcap_{t\in\mathbb{N}} [p_1^t, p_2^t] \neq \emptyset$$

The above intersection is an interval which may be a single point or not, but in any case the sequence of midpoints, that is, $\{trust_d^t\}_{t\in\mathbb{N}}$ converges.

Conclusions and Future Work

We have proposed a trust measure in semantic P2P systems which is based on an estimation of the probability that peers return satisfactory answers. It exploits the information that is included in peers' ontologies and alignments, along with the information that comes from peers' experience.

We presented the computation of this trust measure as more queries are sent and answers received and we have proven the convergence of this computation.

Many probabilistic approaches for trust exist in the field of multiagent systems (Sabater and Sierra 2005). The one of (Mui et al. 2001) also uses a Bayesian approximation but relies on ratings instead of ontology content. One significant difference between these approaches and ours is that we do not assume any malicious behavior: unsatisfactory answers are the result of peers' incapacity to understand each other.

EigenTrust (Kamvar, Schlosser, and Garcia-Molina 2003) is a peer-to-peer algorithm which, like ours, has a direct trust computation. However, it is global in two distinct ways: (a) it computes a unique reputation for each peer, and (b) this value applies to any query while we remain query dependent. Moreover, our direct trust is estimated from the content of local ontologies unlike EigenTrust.

As future work we want to look into oracles, i.e., external ways to evaluate the validity of a answer. We also would like to investigate the trade-offs among ontology, alignment, and query language expressiveness and the constraints that we put on the ontology updating. The combination of trust and information gain will be taken into consideration too.

References

Adjiman, P.; Chatalic, P.; Goasdoué, F.; Rousset, M.-C.; and Simon, L. 2006. Distributed reasoning in a peer-to-peer setting: application to the Semantic Web. *Journal of Artificial Intelligence Research* 25:269–314.

Artz, D., and Gil, Y. 2007. A survey of trust in computer science and the semantic web. *Journal of Web Semantics* 5(2):58–71.

Berners-Lee, T. 1997. Cleaning up the user interface. http://www.w3.org/DesignIssues/UI.html.

Euzenat, J., and Shvaiko, P. 2007. *Ontology matching*. Springer.

Euzenat, J. 2008. Algebras of ontology alignment relations. In *Proc. 7th ISWC*, 387–402.

Halevy, A.; Ives, Z.; Tatarinov, I.; and Mork, P. 2003. Piazza: data management infrastructure for Semantic Web applications. In *Proc. 12th WWW Conference*, 556–567.

Kamvar, S. D.; Schlosser, M. T.; and Garcia-Molina, H. 2003. The EigenTrust algorithm for reputation management in P2P networks. In *Proc. 12th WWW Conference*, 640–651.

Lukasiewicz, T., and Straccia, U. 2008. Managing uncertainty and vagueness in description logics for the Semantic Web. *Journal of Web Semantics* 6(4):291–308.

Mui, L.; Mohtashemi, M.; Ang, C.; Szolovits, P.; and Halberstadt, A. 2001. Ratings in distributed systems: A bayesian approach. In *Proc. 11th WITS*.

Nejdl, W.; Wolf, B.; Qu, C.; Decker, S.; Sintek, M.; and al. 2002. EDUTELLA: a P2P networking infrastructure based on RDF. In *Proc. 11th WWW Conference*, 604–615.

Sabater, J., and Sierra, C. 2005. Review on computational trust and reputation models. *AI Review* 24(1):33–60.

LIG