An Efficient Ontology-Based Expert Peering System

Tansu Alpcan C. Bauckhage S. Agarwal

Deutsche Telekom Laboratories

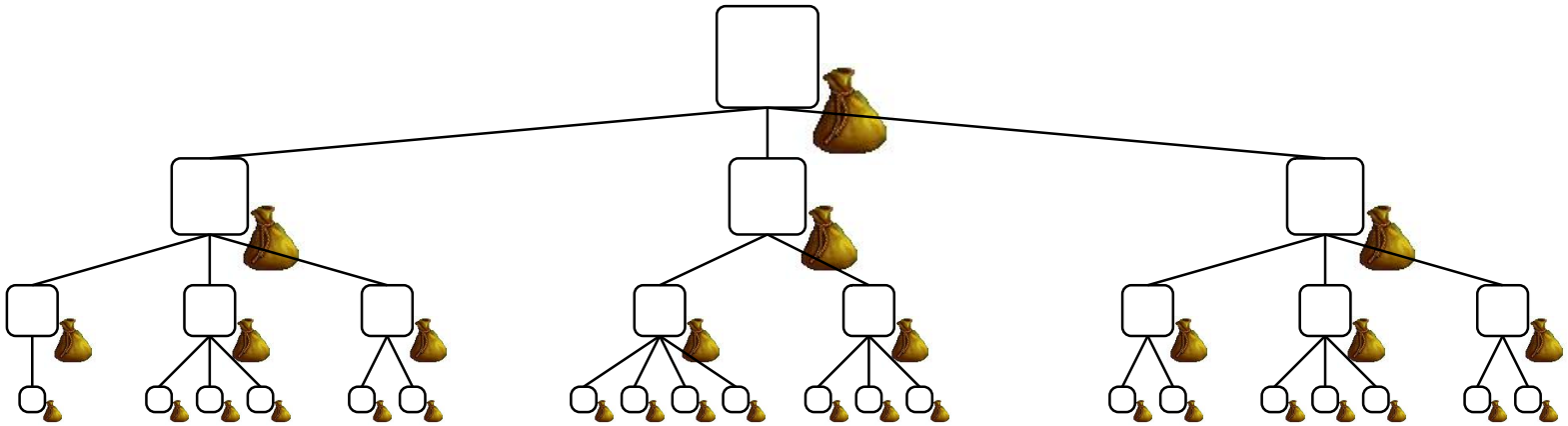
GBR 2007

Outline

Overview

- A novel expert peering system for community-based information exchange
- A graph-based scheme consisting of a taxonomy where each node represents a (sub)topic.
- User queries and profiles of the participating experts are mapped to subtrees of this ontology.
- Assigning queries to relevant experts becomes a problem of graph matching.
- A serialization of the taxonomy allows use of simple dot products on the ontology vector space effectively to address this problem.

The Ontology

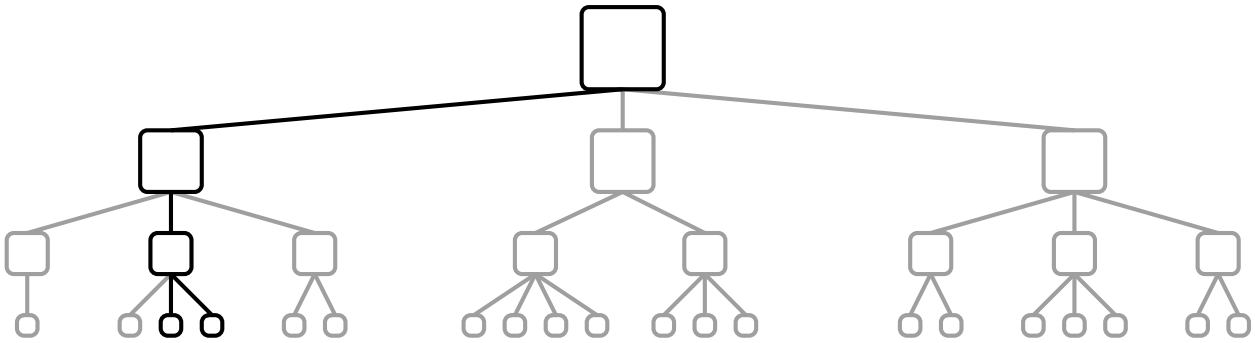


Mapping to Ontology-space

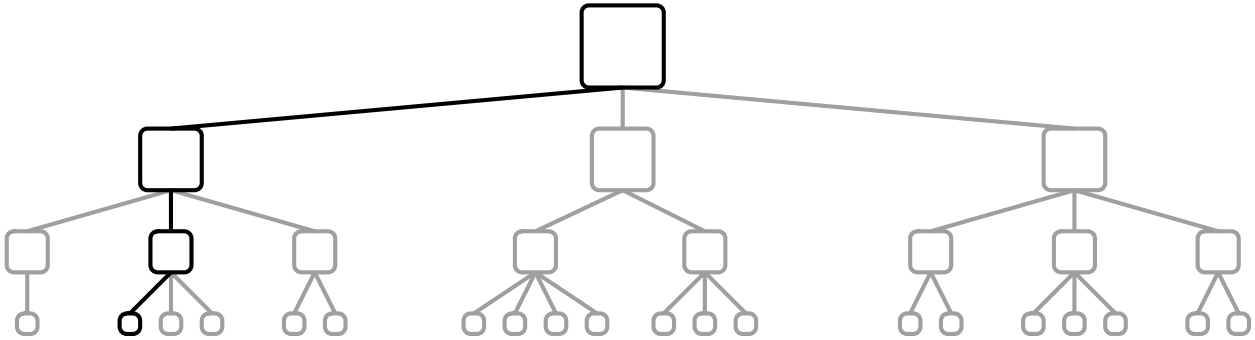
- The ontology tree is serialized to obtain an associated vector representation $\mathbf{v}(T) \in \mathbb{R}^N$.
- Represent entities as subtrees of the ontology and equivalently as vectors on the so called ontology-space $S(T) \subset \mathbb{R}^N$.
- Define a similarity measure $r(i,j)$ between two entities i and j :

$$r(i,j) := \frac{|B(i) \cap B(j)|}{\sqrt{|B(i)|}\sqrt{|B(j)|}},$$

Mapping to Ontology-space



(a) An expert on two topics represented by leaves of the ontology can be modeled as a subtree.



(b) A query on a topic represented by a leave of the ontology can be modeled as a subtree.

Methods Properties

- The ontology (vector) space has a much smaller dimension than the commonly used term-by-document spaces. It also avoids the need for maintaining large, inefficient, and static dictionaries.
- Each dimension of the ontology-space, which actually corresponds to a node (subject), has inherent semantic relations with other nodes.
- One such relation is hierarchical and immediately follows from the tree structure of the ontology.
- It is also possible to define other graph theoretic relations, for example, by defining overlay graphs.

Experiment Setup

- We choose a 245 node subset of an ontology prepared by an UK education agency.
- The procedure for populating the ontology:
  1. The node's name is used in finding 10 top ranked documents through *Yahoo!* search web services.
  2. The obtained HTML documents are converted to a single text document.
  3. This resulting document is further processed using the Natural Language Toolkit (NLTK) by (a) tokenizing, (b) stop word removal, and (c) stemming with Porter's stemmer.
- Random expert and query generation

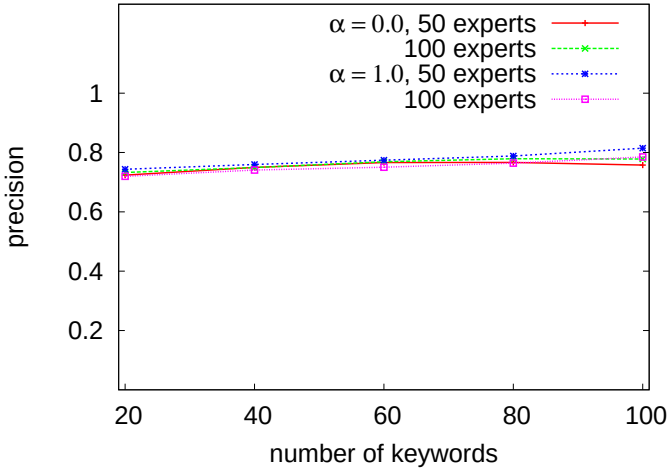
Experiment Setup

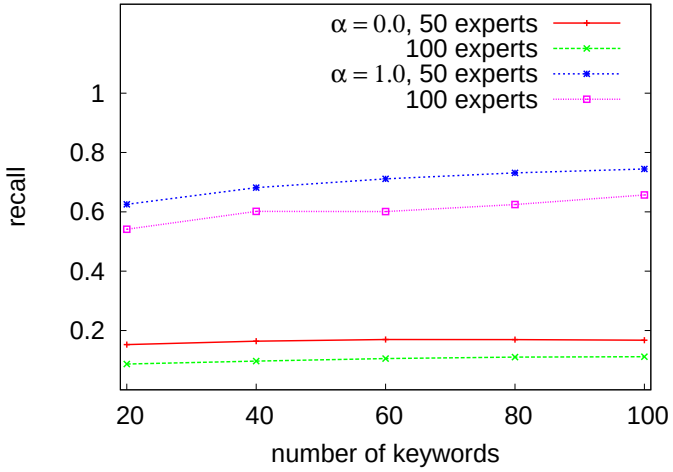
- The expert peering: set of experts $R(q)$ with highest matching score, $m(q,e) := \mathbf{v}(q) \cdot \mathbf{v}(e)$ given query q.
- The "ground truth" vectors are used to calculate the set of "correct" experts $A(q)$
- The recall and precision measures are calculated as the average of $N = 1000$ such queries

$$recall = \frac{1}{N} \sum_{i=1}^{N} \frac{\mid A(q_i) \cap R(q_i) \mid}{\mid A(q_i) \mid}$$

$$precision = \frac{1}{N} \sum_{i=1}^{N} \frac{\mid A(q_i) \cap R(q_i) \mid}{\mid R(q_i) \mid}$$

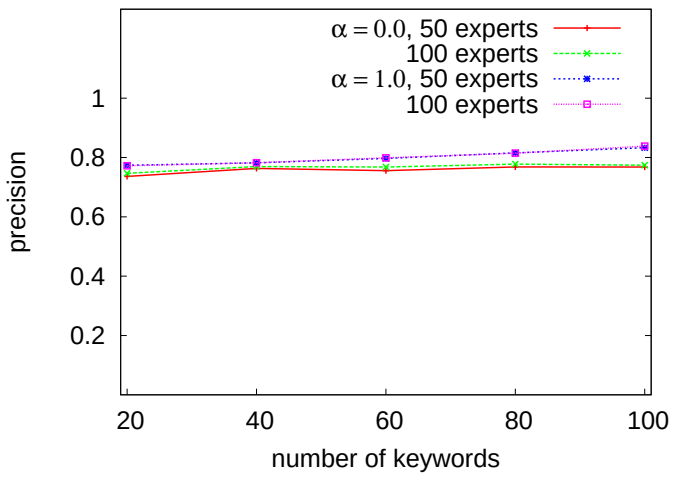
Numerical Results

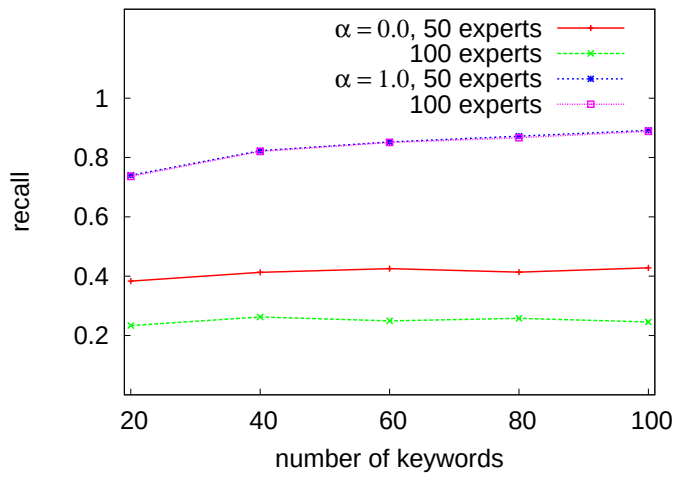




Precision and recall when assigning the set of experts with the top three rankings to each query.

Numerical Results



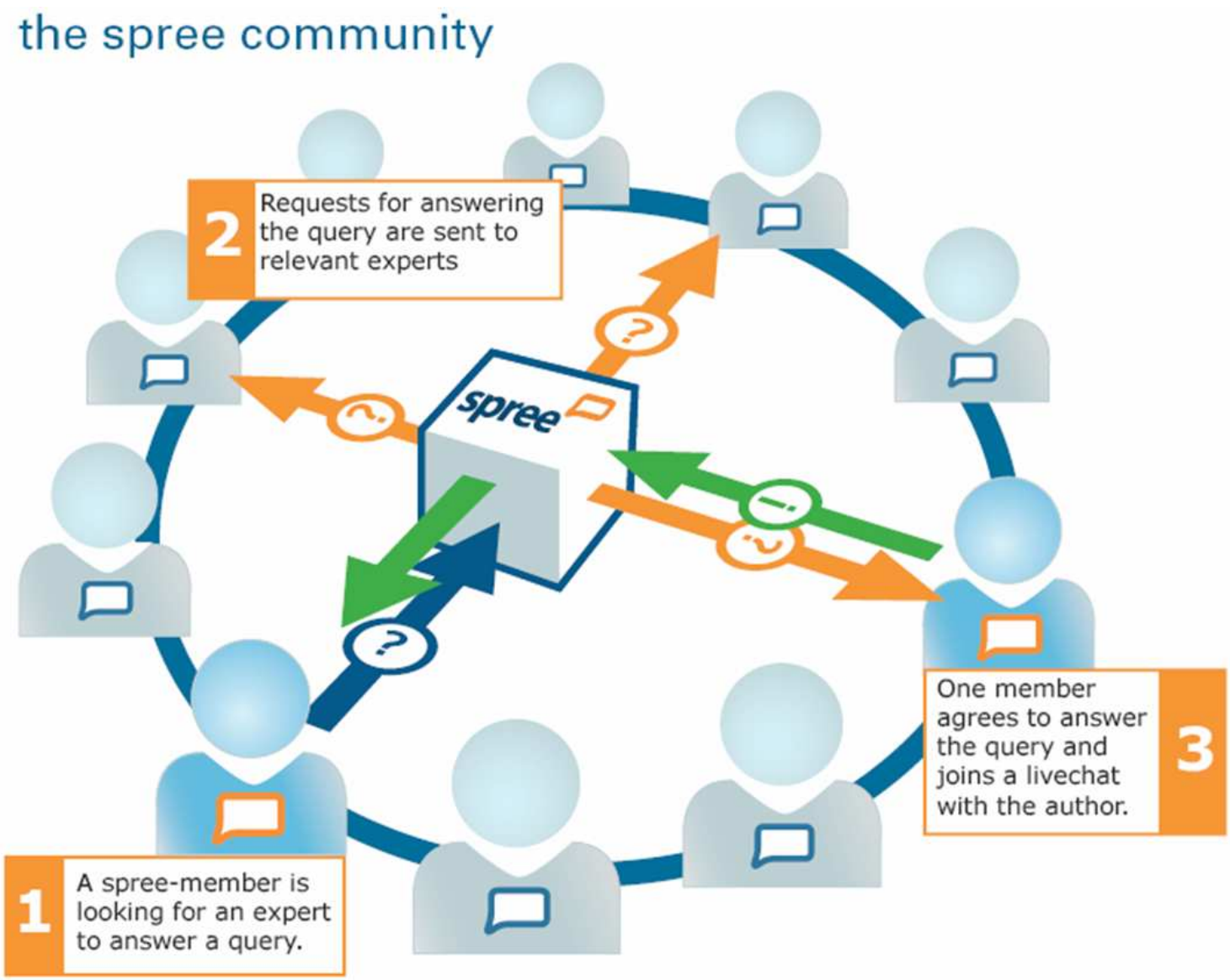


Precision and recall when assigning the set of experts with the top six rankings to each query.

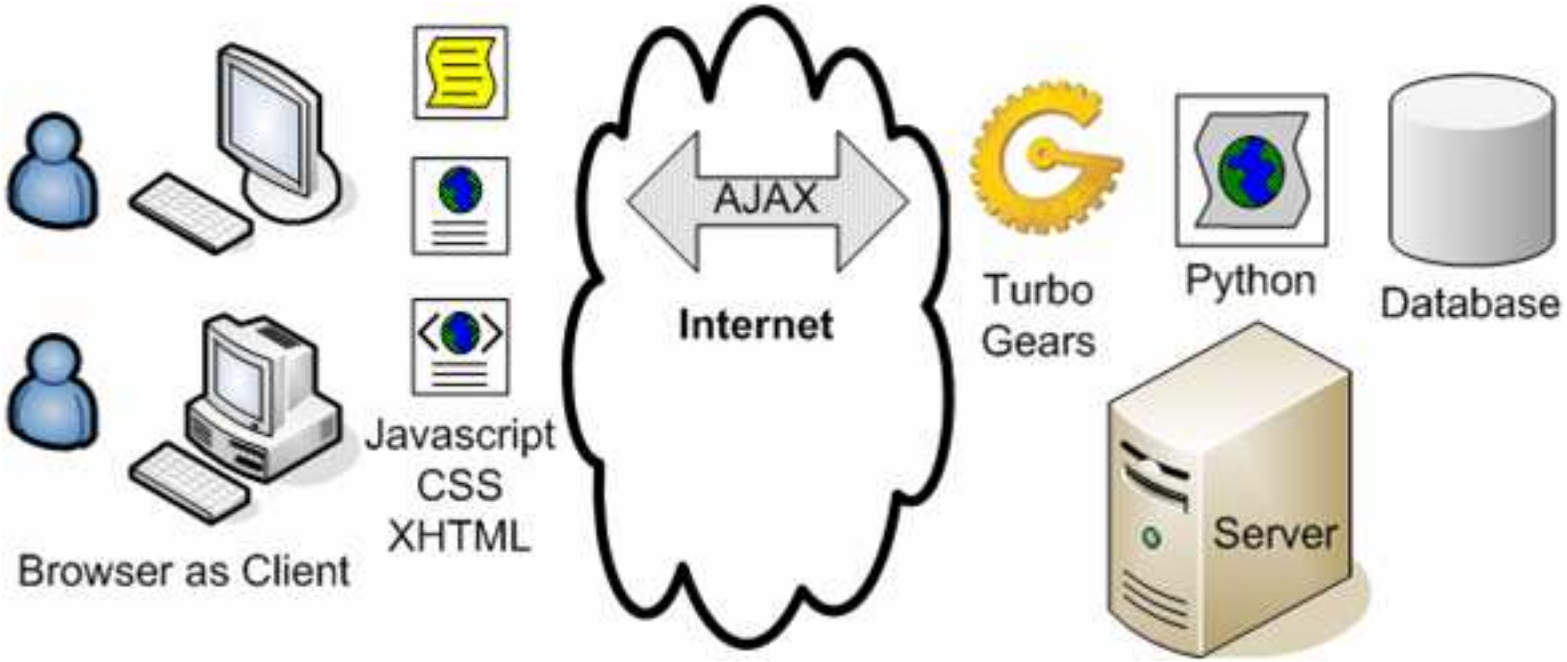
Observations

1. Choosing the larger $\alpha = 1$ value for the algorithm leads to improved results as it restricts unnecessary branching, and hence noise.
2. The precision remains high regardless of the number of experts and α. We attribute this result to hierarchical structure and robustness of our system.
3. With the correct set of parameters we observe that both the precision and recall are relatively insensitive to the number of experts which indicates scalability.
4. The precision and recall curves are rather flat demonstrating that our system performs well in peering the experts even when given limited information.

Application



Architecture



Conclusion

- presented an ontology-based approach for an expert peering and search system.
- described a graph-based representation scheme consisting of an ontology tree where each node corresponds to a (sub)topic and is associated with a bag of words.
- addressed the graph matching problem of assigning queries to relevant experts on a *vector space*, which follows from a serialization of the ontology tree.
- Preliminary experiments demonstrate the efficiency, robustness, and high performance of our algorithm over a range of parameters.
- A prototype system is under development and user testing.

Thank you!

This publication and more information about the Spree project are available on my website:

http://decision.csl.uiuc.edu/~alpcan/

or

http://deutsche-telekom-laboratories.de/~alpcan/