Ontology Learning from Text

Discovering Non-Taxonomic Conceptual Relations from Text

http://www.aifb.uni-karlsruhe.de/WBS/

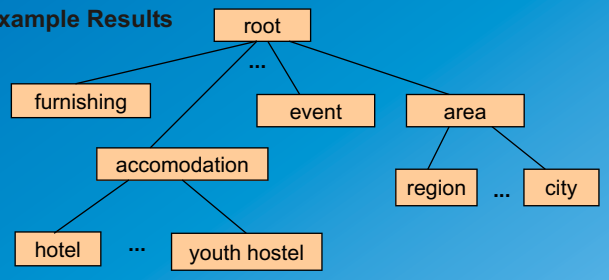
Approach

New approach for discovering non-taxonomic conceptual relationships from natural language texts building on shallow text processing techniques.

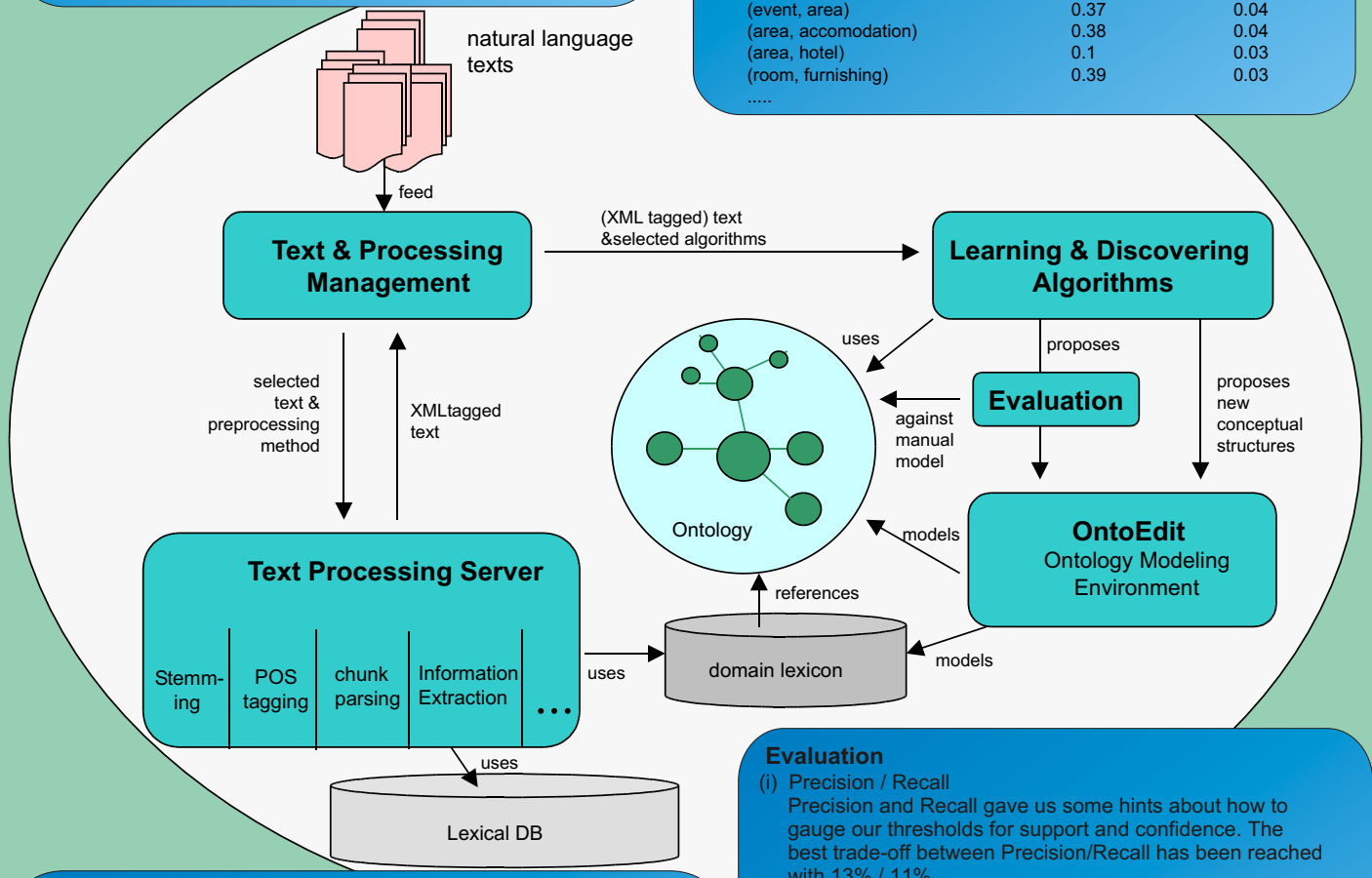
We use a generalized association rule algorithm that does detect relations between concepts and determines the appropriate level of generalization.

An empirical evaluation is performed against a hand-modeled ontology in the tourism domain. We have performed the evaluation using standard measures (precision, recall) and have also conceived a new evaluation measure, RLA, to take account of the sliding scale of adquacy prelevant in a hierarchical structure.

Example Results



Relation	Confidence	Support
(costs, accomodation)	0.38	0.04
(event, area)	0.37	0.04
(area, accomodation)	0.38	0.04
(area, hotel)	0.1	0.03
(room, furnishing)	0.39	0.03
.....		



Background

Ontologies have become an important means for structuring information and information systems.They have shown their usefulness in application areas such as **information integration**, **information brokering** and **natural language processing.**

There remains the problem of engineering large and adequate ontologies within short time frames. Ontology Engineering is a complex and time intensive task. It is very difficult and cumbersome to manually derive ontologies from existing data (e.g. a corporate intranet).

Ontologies consist of a set of concepts, which are taxonomically and non-taxonomically related by object relations. On the basis of the object model a set of logical axioms enforce semantic constraints. Non-taxonomic conceptual relations between concepts appear as a major building block in common ontology definitions.

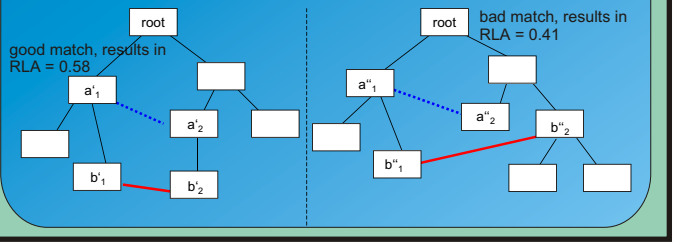
Evaluation

(i) Precision / Recall

Precision and Recall gave us some hints about how to gauge our thresholds for support and confidence. The best trade-off between Precision/Recall has been reached with 13% / 11%.

(ii) Relation Learning Accuracy (RLA)

RLA evaluates the degree to which relations that are learned match existing relations. The max. RLA with 0.69 is reached with support 0.04 and confidence 0.1. The RLA compaired against a uniform distribution of all possible relations results in 0.39, and thus, is significantly worse than learning results.



AIFB

Knowledge Management Group
Institute AIFB - Karlsruhe University
Alexander Maedche & Steffen Staab
{maedche, staab}@aifb.uni-karlsruhe.de