Ontological Representations for Supporting Learning in Business Communities

Bojan Cestnik
Temida d.o.o. & Jozef Stefan Institute, Ljubljana, Slovenia
bojan.cestnik@temida.si

Tanja Urbančič
University of Nova Gorica, Slovenia
tanja.urbancic@ung.si

Ingrid Petrič
University of Nova Gorica, Slovenia
tanja.urbancic@ung.si

Presented by

Stojan Košti
Temida d.o.o., Ljubljana, Slovenia
stojan.kosti@temida.si

Contents

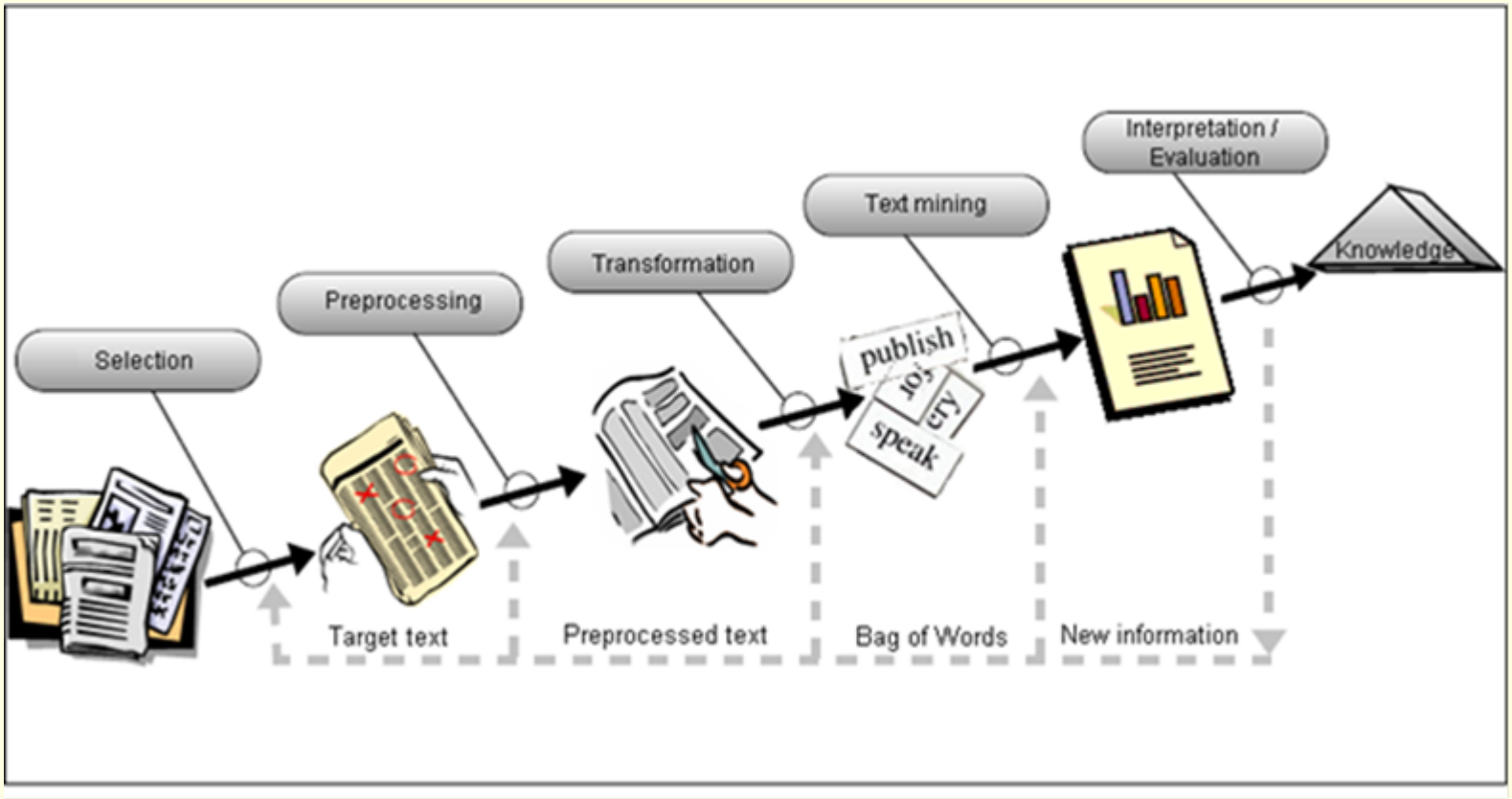
- Introduction
- Background
- Ontology construction case
- Evaluation of results
- Conclusions

Introduction

- Learning tasks in business communities
  - Adequately recognize partners‘ products and skills
  - Explore heterogeneous data sources
  - Modern techniques (data mining, text mining, etc.)
- Knowledge acquisition from various sources
- Business knowledge representation: topic ontology

Extracting knowledge from text

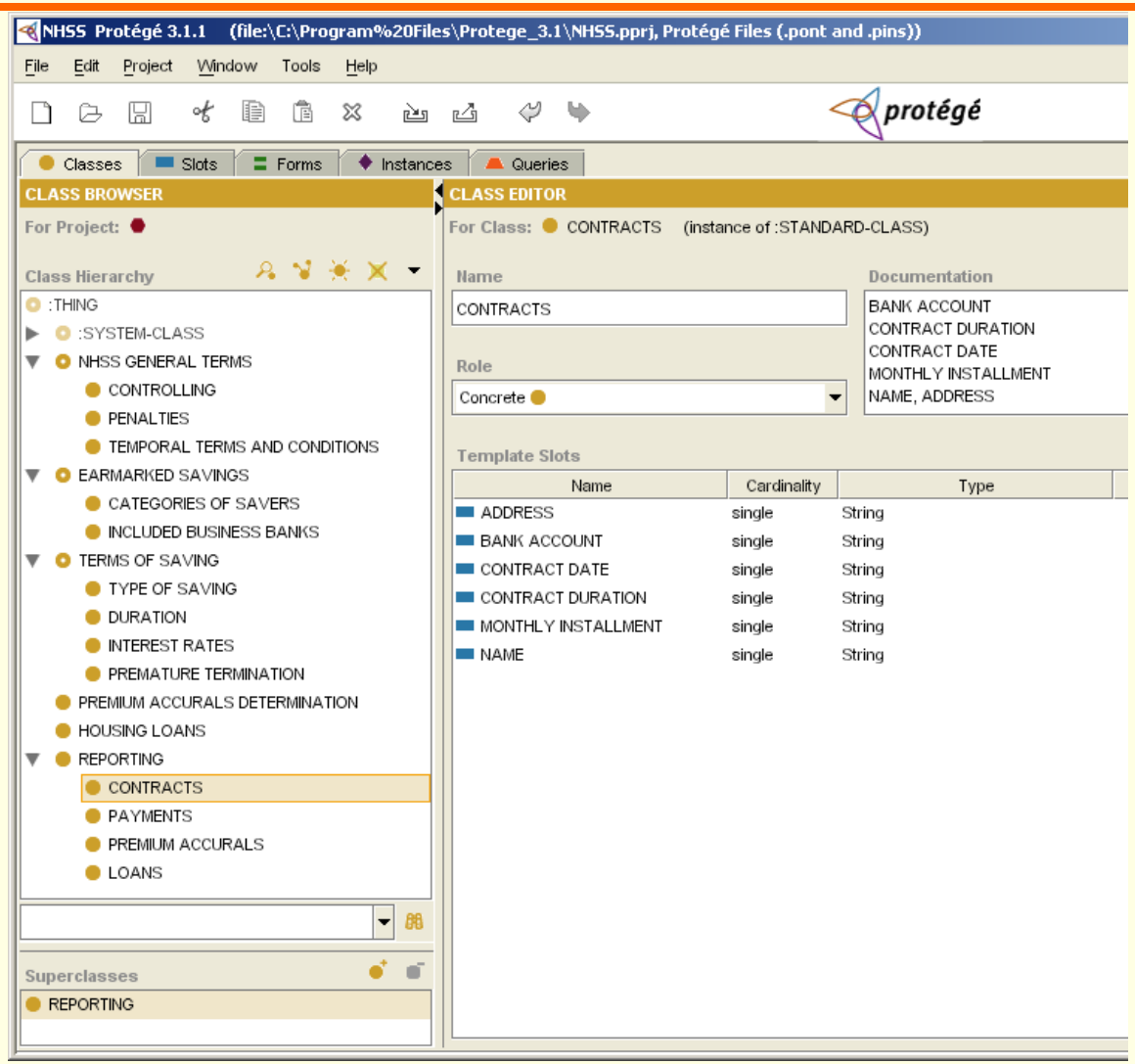
- Text mining steps:



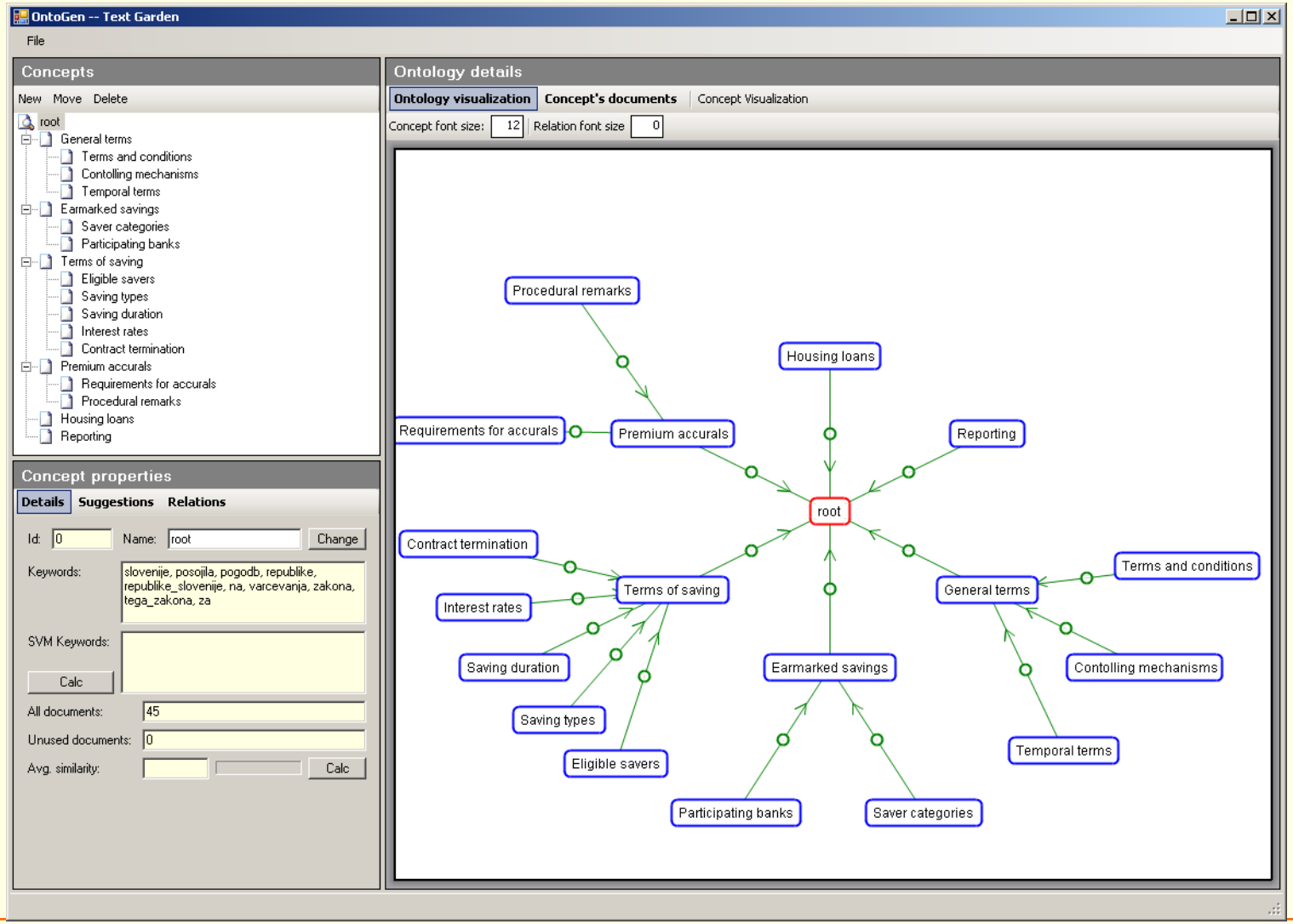
Introduction to ontologies

- Ontologies
  - Organize scientific information
  - Common vocabulary of concepts
  - Contribute to common understanding of problem domains
- Formal representation of a shared conceptualization of a particular domain
- Construction of ontologies
  - Manual (e.g. Protégé)
  - Semi-automatic (e.g. OntoGen)

NHSS *Protégé* ontology



NHSS *OntoGen* ontology



Ontology learning from text

- Constituting parts of ontology
  - Terms — *tender, appartment, buyer*
  - Synonyms — *appartment, flat, dwelling*
  - Concepts — *appartment:={1-room, 2-room, …}*
  - Taxonomy — *is_a(buyer, person)*
  - Relations — *buy(buyer, appartment)*
  - Rules — $\forall x, y : (has\_a\_child(x, y) \rightarrow family(x))$
- Methods
  - Statistical
  - Linguistic
  - Hybrid

Areas of text processing

- Information retrieval – Search in DB
- Natural language processing – Computational linguistics
- Text mining – Data analysis
- Semantic web – Knowledge representation and Reasoning

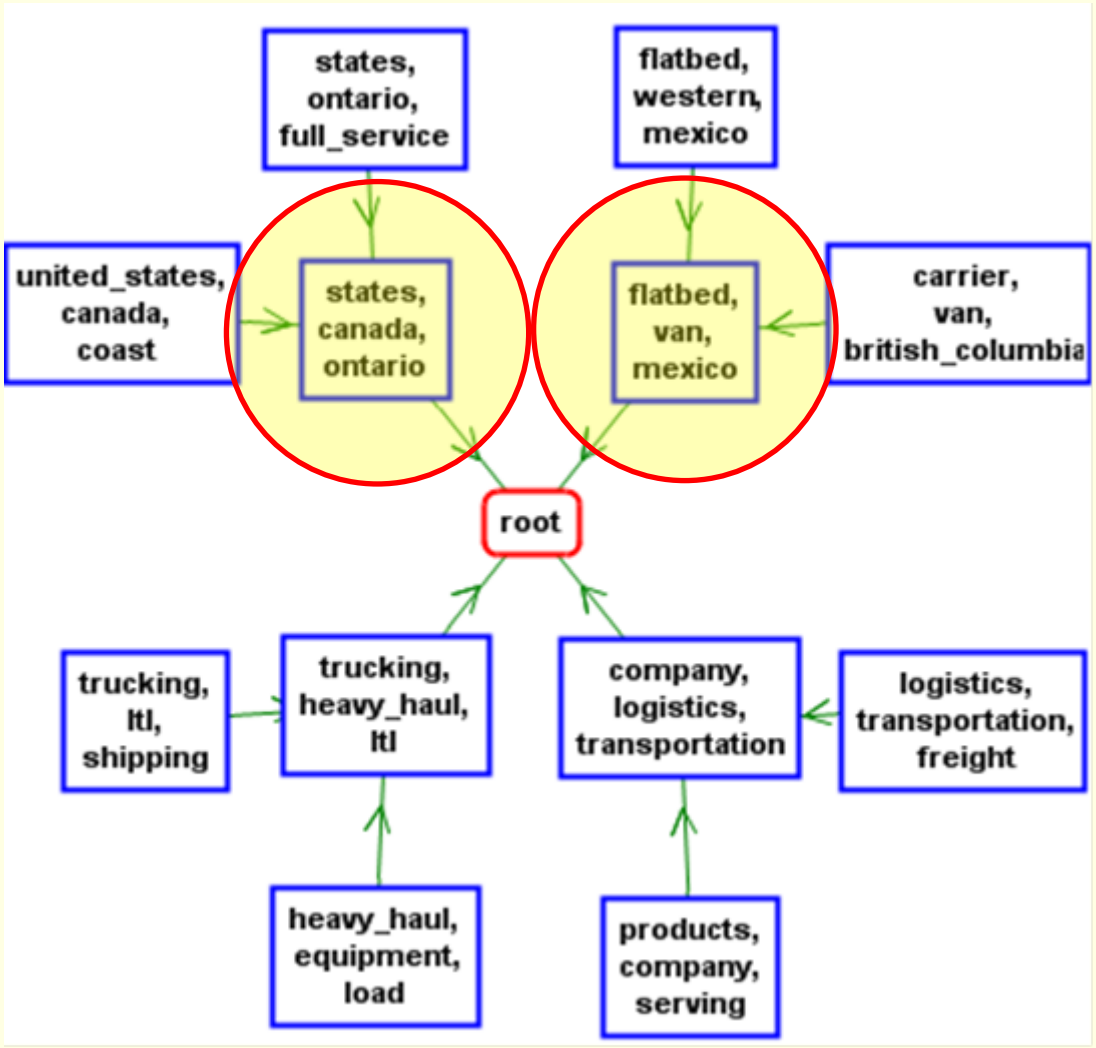
Elements of text processing

- Word
- Sentence
- Paragraph
- Document
- Collection of documents

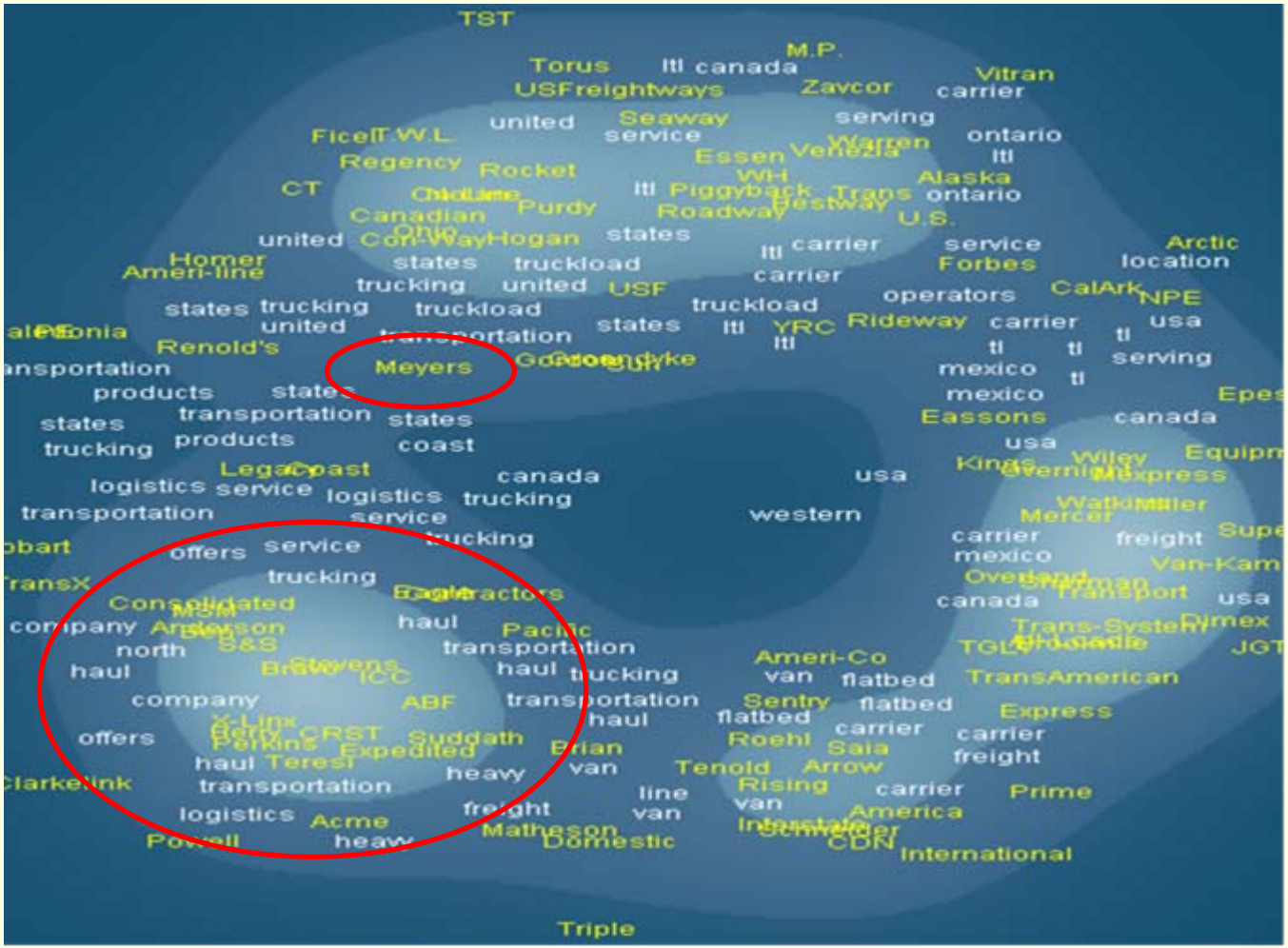
Ontology construction case

- Input document formats: XML, HTML, doc, pdf
- Preprocessing
  - Removing graphics, markup and meta tags
  - Lemmatization
  - Stop-words removal
- Tool: OntoGen
- Case: North American transportation domain
- Input documents
  - Professional descriptions of NA transport companies
  - From Open Directory Project DMOZ

Topic ontology: cluster descriptions



Document map



company name

keyword

cluster

Conclusion I

- Ontologies for efficient learning in business communities
- OntoGen: constructing ontology for a given problem domain from description documents
  - Obtain overview of a given problem domain
  - Requires substantially less time to construct
  - Can be useful for exploiting existing but overlooked knowledge

Conclusion II

- Constructed ontology quickly reveals top level concepts
  - Birds-eye view over a given problem domain
  - Result: easier and more effective business concepts identification and management