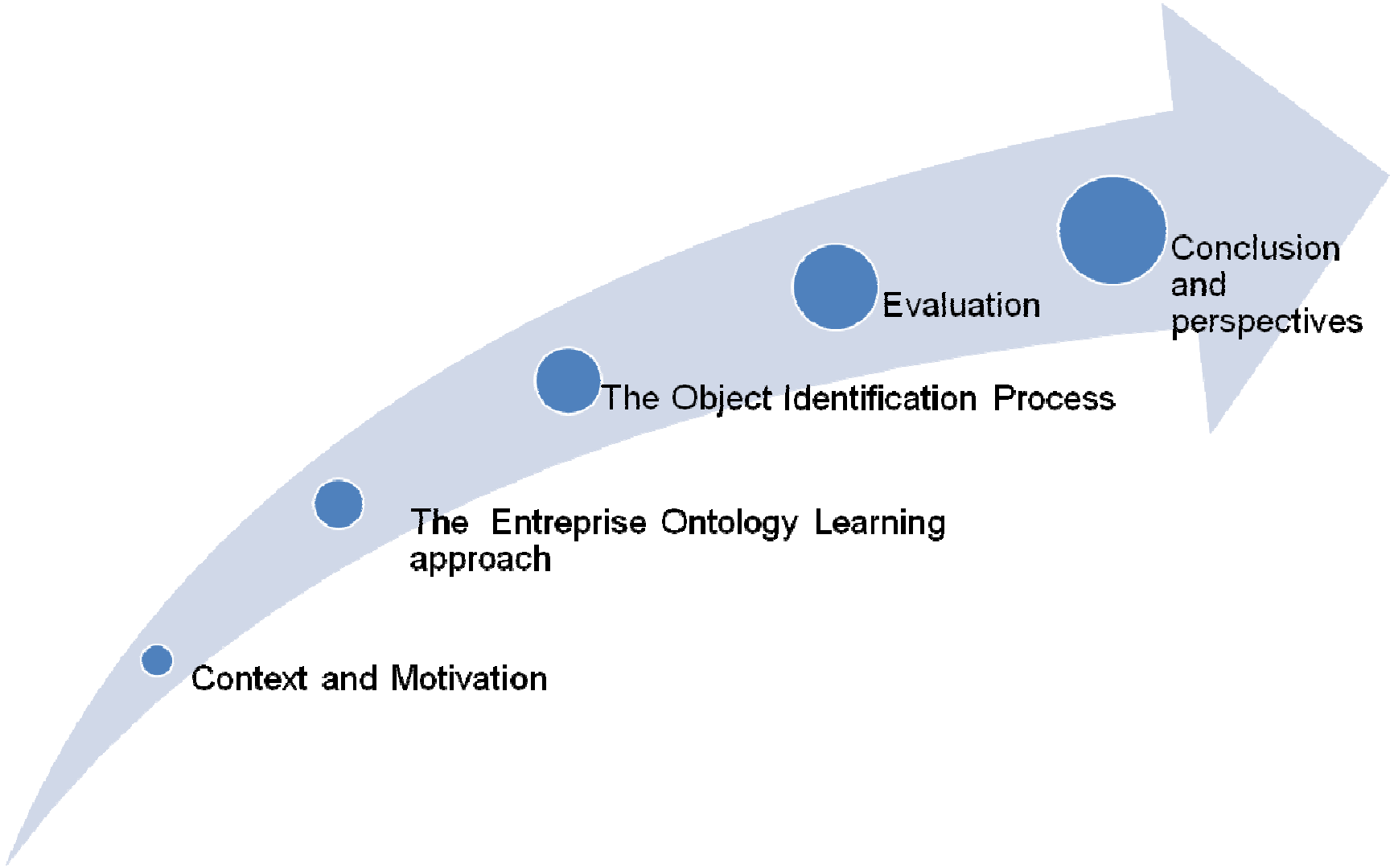
Enterprise Ontology Learning for Heterogeneous Graphs Extraction

Rania Soussi & Marie-Aude Aufaure
Laboratoire MAS – Ecole Centrale Paris

03/12/2012

Outline


Context and Motivation
The Entreprise Ontology Learning approach
The Object Identification Process
Evaluation
Conclusion and perspectives

Context and Motivation

People need to visualize different types of interactions between heterogeneous objects: products and locations, customers and products, social network...).

Graphs are a structure relevant to analyze these interactions and facilitate their querying

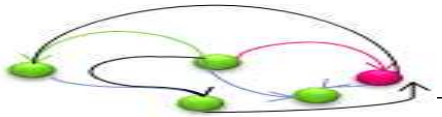
Graph Model Extraction

Relational database

Object Interaction Graph Extraction

Object Identification

Relation Extraction



Need for external Resources

Use of Enterprise Ontology

Context and Motivation

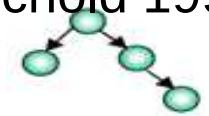
- Enterprise ontology is an ontology used to describe the domain, or parts of the domain of an enterprise [Blomqvist 2007].
- The proposed approaches of enterprise ontology building are too general or too specific.

Need for a new approach combining the two types of concepts.

The Entreprise Ontology Learning Approach:
Overview

Unstructured Documents

[Uschold 1995]



Existing Enterprise Ontologies

Enterprise Pattern and Model

[Silverston 2001]

Building the Generic Ontology

Specific Enterprise Ontology Learning

Objective : build an ontology that contains generic concepts of the domain and concepts related to a specific enterprise

The "Entreprise Ontology Learning" Approach

Building the Generic Part-1-

- The approach uses external ressources:
  - Existing ontologies [Uschold 1995]
  - Patterns from other domains [Silverston 2001]

- The methodology used is similar to the one proposed by [Noy, 2001]
  - Regroup the common concepts
  - Define the relations between the concepts: Synonyms, hierarchical,…
  - Collects the concepts attributes

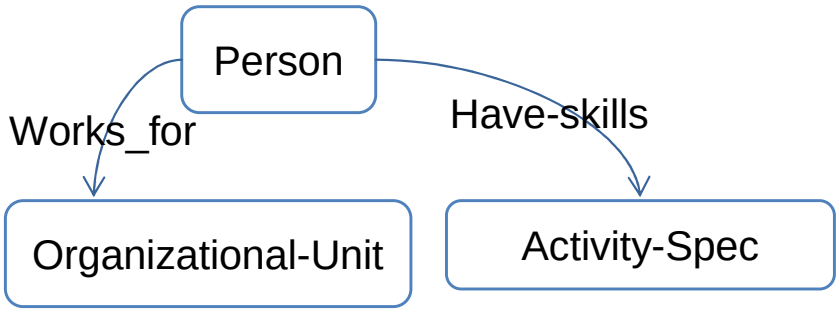
The "Entreprise Ontology Learning Approach

Building the Generic Part-2-

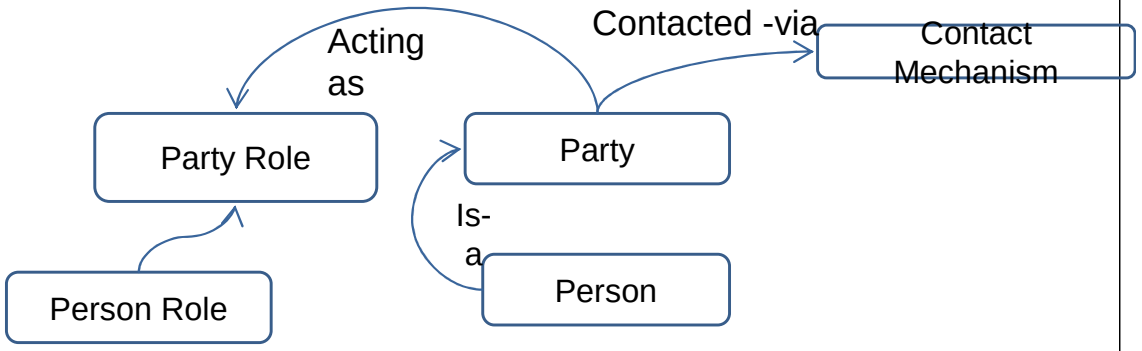
Step1:
Regroup commun concepts

Step2:
Define relations

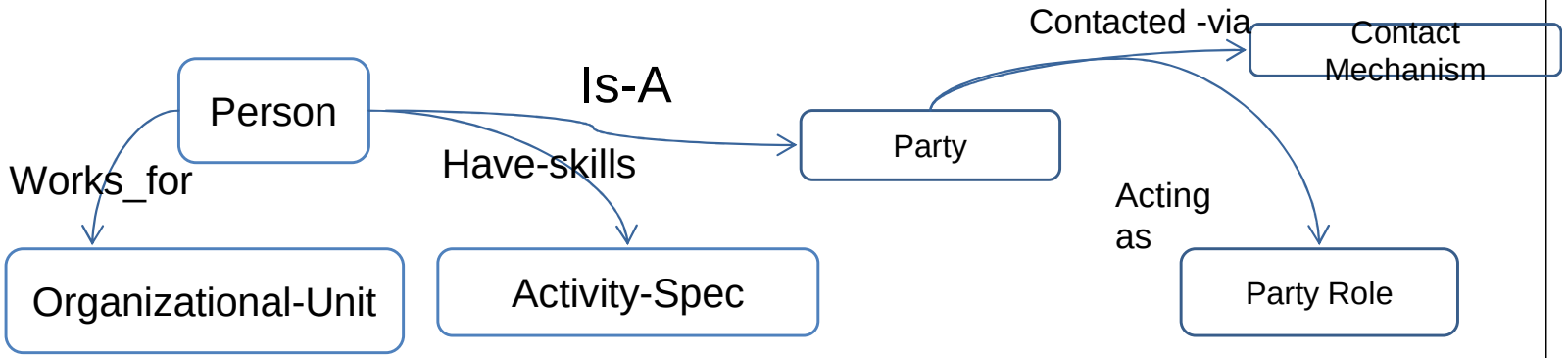
Step3:
Specify concepts properties



Ushold ontology



Silverston data model



The Enterprise Ontology Learning Approach

Building the specific part

Input: The generic ontology

Step1:Document Treatment

The documents are annotated using linguistic and semantic analysis

Step2: Learning process

2.1 candidate elements are detected using Lexico-Syntactic patterns.

CR:=<t,ce,c,r>

2.2. Candidate elements are analyzed

WebOverlap(c,ce):= $\frac{hits(c\ and\ ce)}{Min(hits(c), hits(ce))} > \alpha$

2.3 New patterns detection

Step3:Population process

These include corporations, cooperatives, sole traders and other organizations.

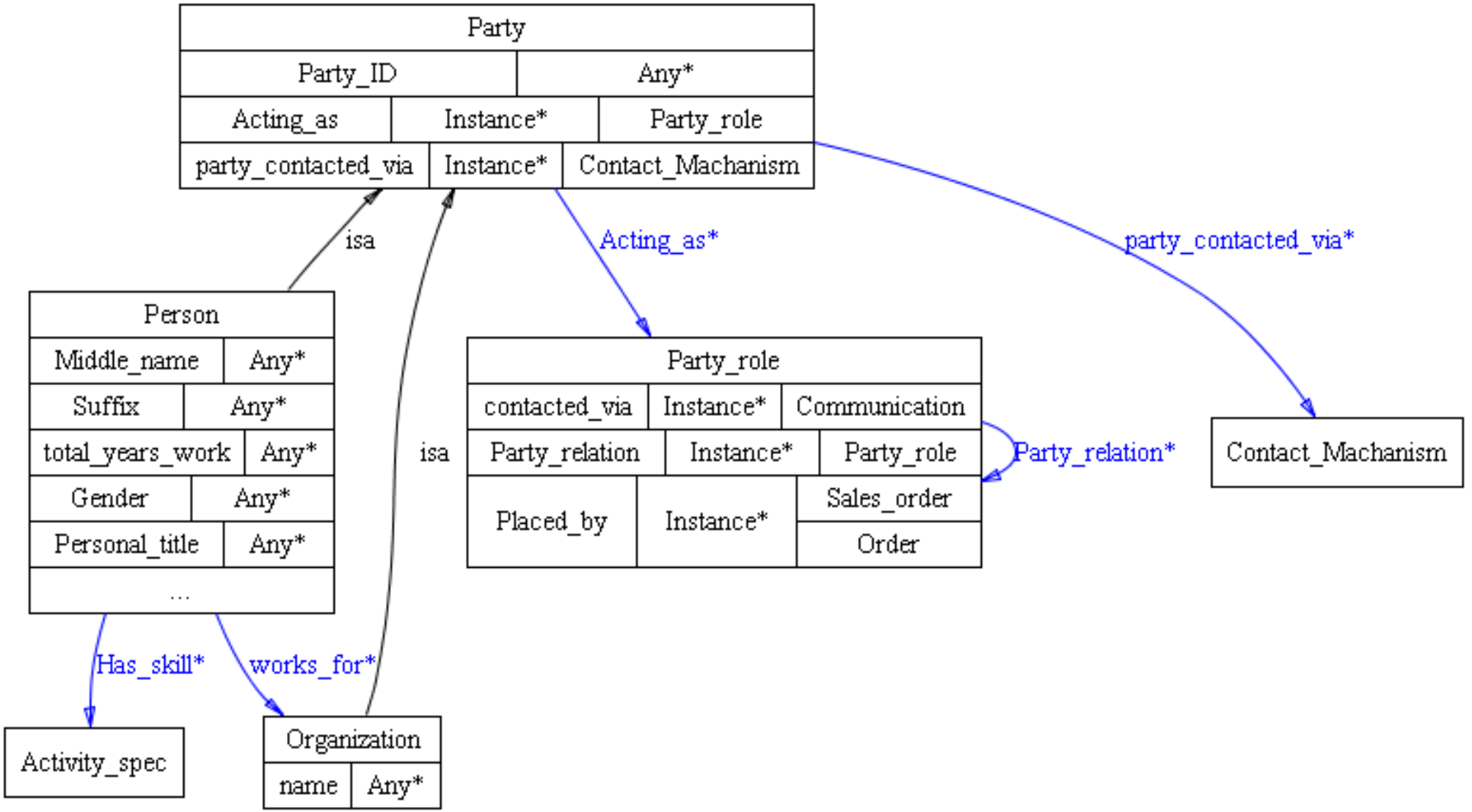
Detect Hearst Pattern
NP {NP,}* {,} {and|or} other

CR1:=<OP, corporation, organization, IS-A>
CR2:=<OP, cooperative, organization, IS-A>
CR :=<OP, sole trader, organization, IS-A>

WebOverlap(corporation, organization)=0,305
WebOverlap(cooperative,organization) =0,499
WebOverlap(sole trader, organization)=0,002

Approach

Part of the enterprise ontology

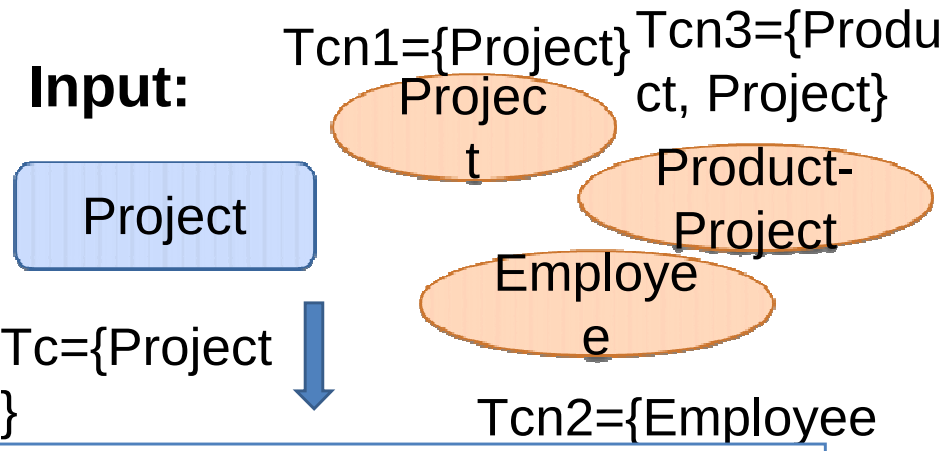


Object Identification Process

- **Step 1. Name Treatment**
  - Extract the Token sets from each node and concept
- **Step 2. Candidate node Extraction**
- Case1: In the case of Tc is (equal|synonym| subconcept) (to|of) Tcn → CN is an object
- Case 2: one of Tc elements are sufix or prefix of at least one of Tcn elements
- Case 3.

$$\frac{\sum_{t_1 \in T_c}\left[\max_{t_2 \in T_{CN}} sim\ (t_1, t_2)\right] + \sum_{t_2 \in T_{CN}}\left[\max_{t_1 \in T_c} sim\ (t_1, t_2)\right]}{|T_c| \times |T_{CN}|}$$

- **Input:**

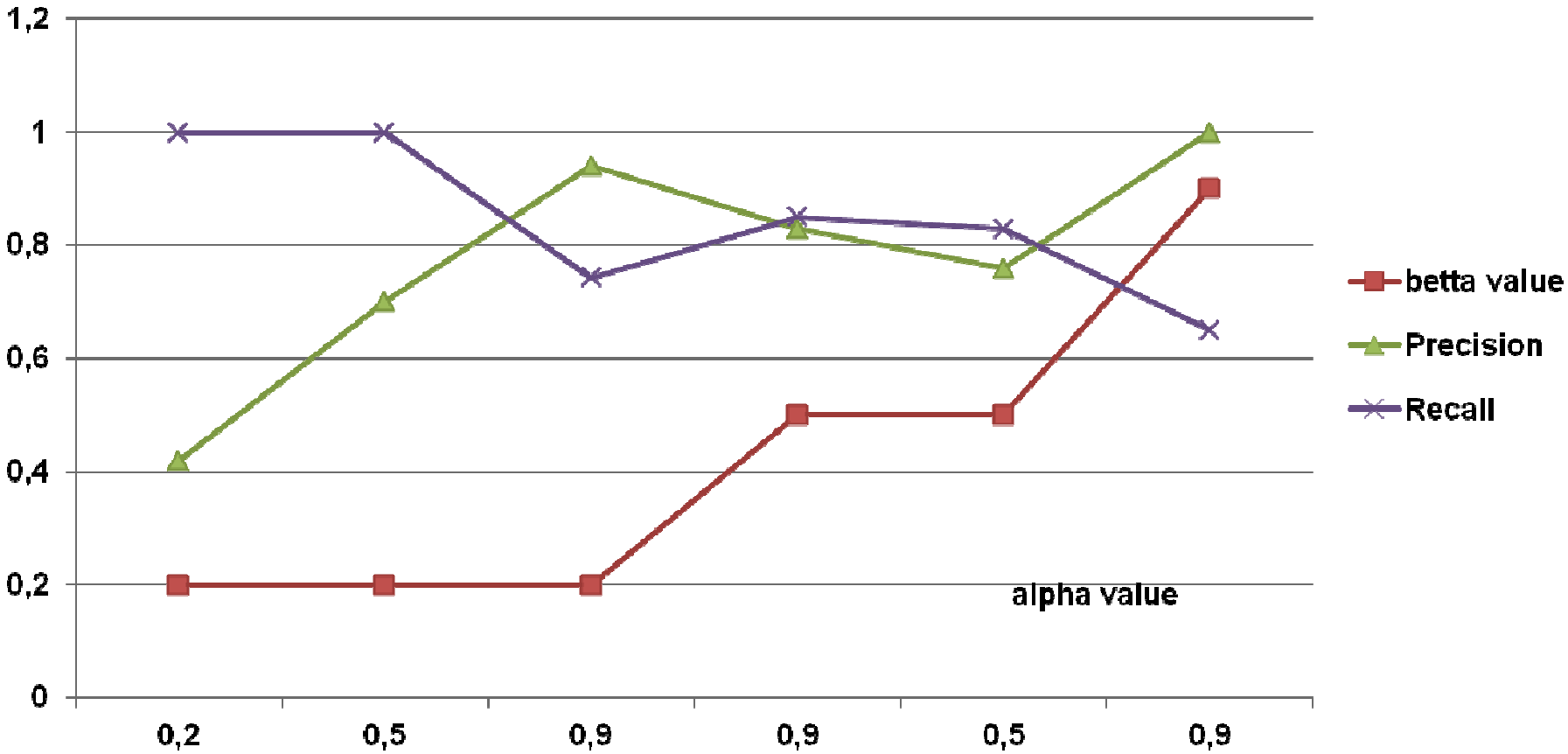


- **Step3. candidate filtring**
- Calculate the attribute similarities

$$Simatt\ (C, CN) = \frac{\sum_{n_c \in DP}\sum_{n_{nc} \in A_{nc}} sim\ (n_c, n_{n_h})}{|DP| + |A_{nc}|}$$

Evaluation

Graph containing 90 heterogeneous objects in the schema level



Conclusion and perspectives

- we have presented an approach for learning Entreprise Ontology from enterprise documents.
- The proposed approach allowed obtaining an ontology containing general concepts of the business domain and specific ones for a particular enterprise.
- The resulting ontology has been used in an approach of interacting objects graph extraction.
- As a future step of our work, we will use the value of the existing instances to improve the matching between the ontology concepts and the graph nodes.
- We will use the ontology relations to enhance the extracted graph.

Questions?