A Fuzzy Ontology-Approach to improve Semantic Information Retrieval

Silvia Calegari

Universita di Milano-Bicocca
Milano, Italy

Elie Sanchez

University of the Mediterranean
Marseille, France

November 12th, 2007, Busan, Korea

Open issues with ontologies

- In ontology-based applications information can often be *ill-defined, uncertain* or *imprecise*

  - For instance, in the semantic-based applications of the Semantic Web (e-commerce, knowledge management, web portals, etc.)

- Humans use *linguistic expressions* to specify their interests and needs

  - A user can be interested in finding *a good asiatic restaurant, a very sunny resort, relevant patents* [there are two kinds of relevance: *query relevance* and *topic relevance*, both are graded concepts]

- Necessity to handle the richness of natural languages used by humans

Formal Fuzzy Ontology Structure

- Defined as a quadruple $\boxed{O_F = (C, R, F, A)}$

- **C** is a set of *Fuzzy Concepts*, or *Entities (**E**)*, indifferently

- **R** is a set of *Fuzzy Relations* in E^n, $R = T \cup T_{not}$, where **T** is for *Taxonomic Relations* and T_{not} for *non-Taxonomic Relations*

- **F** is a set of *Fuzzy Relations*, $F \in \mathbf{F}$ is a relation $F : E^{n-1} \times P \rightarrow [0,1]$, where P is a set of integers, strings, …

- **A** is a set of *Axioms* expressed in a proprer logical language, i.e. predicates that constrain the meaning of concepts, relations, functions

Fuzzy Knowledge Base

- Defined as a couple $\boxed{\mathbf{KB} := (\mathbf{O}_F , \mathbf{I})}$

  where $\mathbf{O}_F$ is a *Fuzzy Ontology*, and $\mathbf{I}$ is a set of associated *Instances*.

  Furthermore, every concept $C \in \mathbf{C}$ is a fuzzy set on $\mathbf{I}$, i.e. $C : \mathbf{I} \to [0,1]$

- In this context, a fuzzy knowledge base is made up of the knowledge

  expressed by the fuzzy ontology and by the set $\mathbf{I}$ of objects stored

  in a data base, i.e. $\mathbf{O}_{DB} = \mathbf{I}$ and $C : \mathbf{O}_{DB} \to [0,1]$.

  In particular $\mathbf{O}_{DB}$ can consist of documents, digital pictures, notes …

Remark

- In a fuzzy knowledge base, instances have been separated from fuzzy ontologies (even an OWL ontology "may" only include instances)

  The advantage is that one can have one ontology and multiple sets of instances that conform to it

Correlations: new fuzzy relationships

- A crucial topic is to define a dynamic knowledge of a domain adapting itself to the context:

- *Trade-off between the correct definition of an object (given by the ontology structure) and the actual meaning assigned to the artifact by users (i.e. the experience-based context assumed by every person according to his specific knowledge)*

- Introduction of *semantic correlations* among the entities that are *searched in a query or when a document has been inserted into a data base*

- Correlations can be defined by a *semantic evaluation of objects* stored in a data base, *query representations* and *ontology structure* (with *fuzzy weights*)

FCN: Fuzzy Concept Network

- A *FCN* is a complete weighted graph $\mathbf{N}_f = (\mathbf{E}, F, m)$

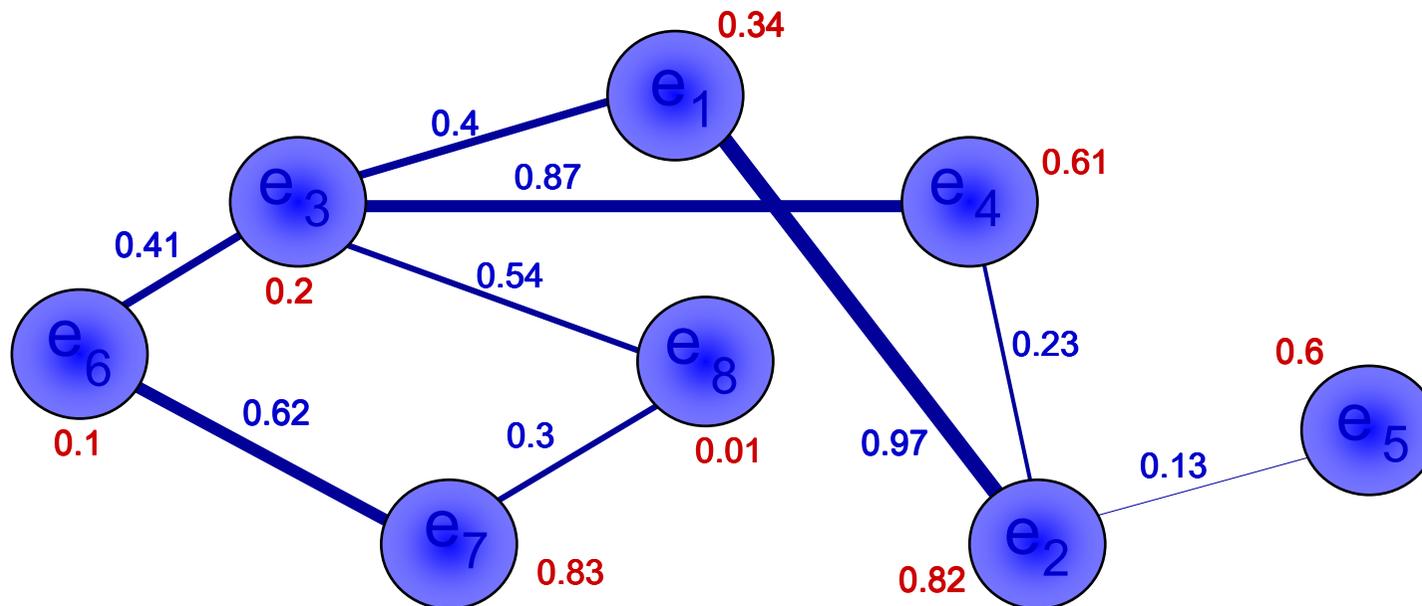 The edges among the nodes in $\mathbf{E}$ are described by the function

  $F: \mathbf{E} \times \mathbf{E} \rightarrow [0,1]$, defining *correlations*

  If $F(e_i, e_j) = 0$ the entities are considered uncorrelated

  Each node e_i is characterized by a membership value defined by

  the function $m: \mathbf{E} \rightarrow [0,1]$, which determines the importance of the

  entity in the ontology

  By definition, $F(e_i, e_i) = m(e_i)$

A Fuzzy Concept Network

$F(e_3, e_8) = 0.54$ (*correlation*)

$m(e_1) = 0.34$

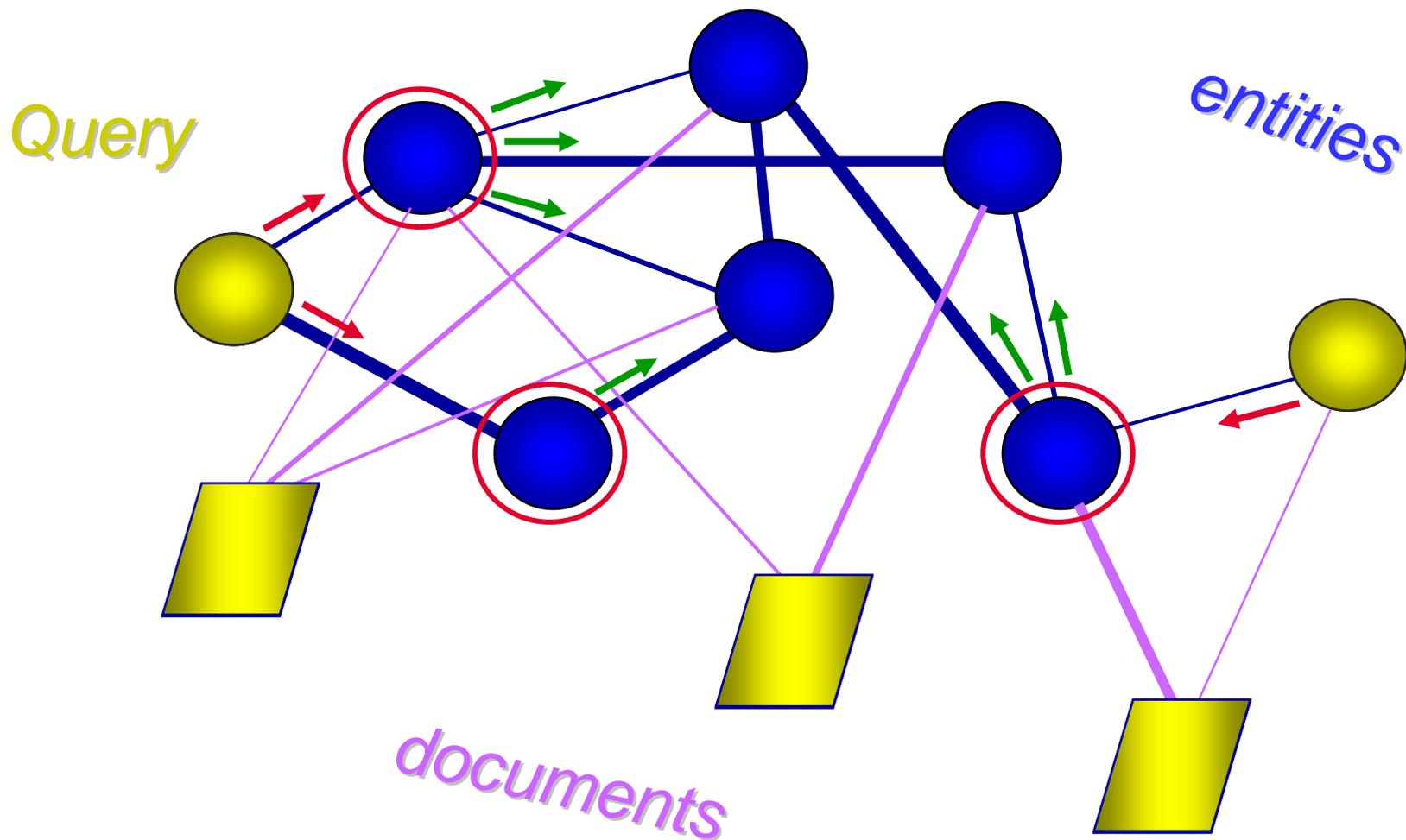# O-FCN: Object-Fuzzy Concept Network

An *O-FCN* is a weighted graph $\mathbf{N}_{f_0} = (\mathbf{O}_{DB} , \mathbf{N}_f)$, where

- $\mathbf{O}_{DB}$ is the set of objects stored in the data base

- $\mathbf{N}_f = (\mathbf{E}, F, m)$ is a *Fuzzy Concept Network (FCN)*

- In $\mathbf{O}_{DB}$, each object o_i is described by the entities of the FCN:

  $o_i = \{ e_1, \ldots, e_n \} \in \mathbf{E}$

Object-Fuzzy Concept Network

Query

entities

documents

- *For each term specified in the query, a unique path is defined at each step, corresponding to the maximum value correlation*

Object-Fuzzy Concept Network

✓ *Introducing* **Fuzzy Ontologies** *in* **Information Retrieval** *allows :*

- *to exploit additional knowledge* **hidden** *in the entities-documents relationships, after querying*

- *to enrich the* **semantics** *of the system*

Fuzzy Ontology-Based Information Retrieval

```
┌──────────────┐  ┌──────────────┐        ┌──────────────────┐
│ Fuzzy Logic  │  │ Ontologies   │        │     Query        │
└──────────────┘  └──────────────┘        │ Representation   │
        ↘            ↙                     └──────────────────┘
      ┌──────────────┐      ┌──────────────────────┐   ↙   ┌──────────────────┐
      │    Fuzzy     │ ───▶ │ Fuzzy Concept Network │ ◀─── │ Database Objects │
      │  Ontologies  │      │       (FCN)          │       └──────────────────┘
      └──────────────┘      └──────────────────────┘
                                    ↘         ↙
                        ┌──────────────────────────────────┐
                        │  Object-Fuzzy Concept Network    │
                        │            (O-FCN)               │
                        └──────────────────────────────────┘
                                        │
                                        ▼
   ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐      ┌──────────────────────────────────┐
   │ Fuzzy Ontology-Based │----▶ │            Semantic              │
   └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘      │     Information Retrieval        │
                                 └──────────────────────────────────┘
```
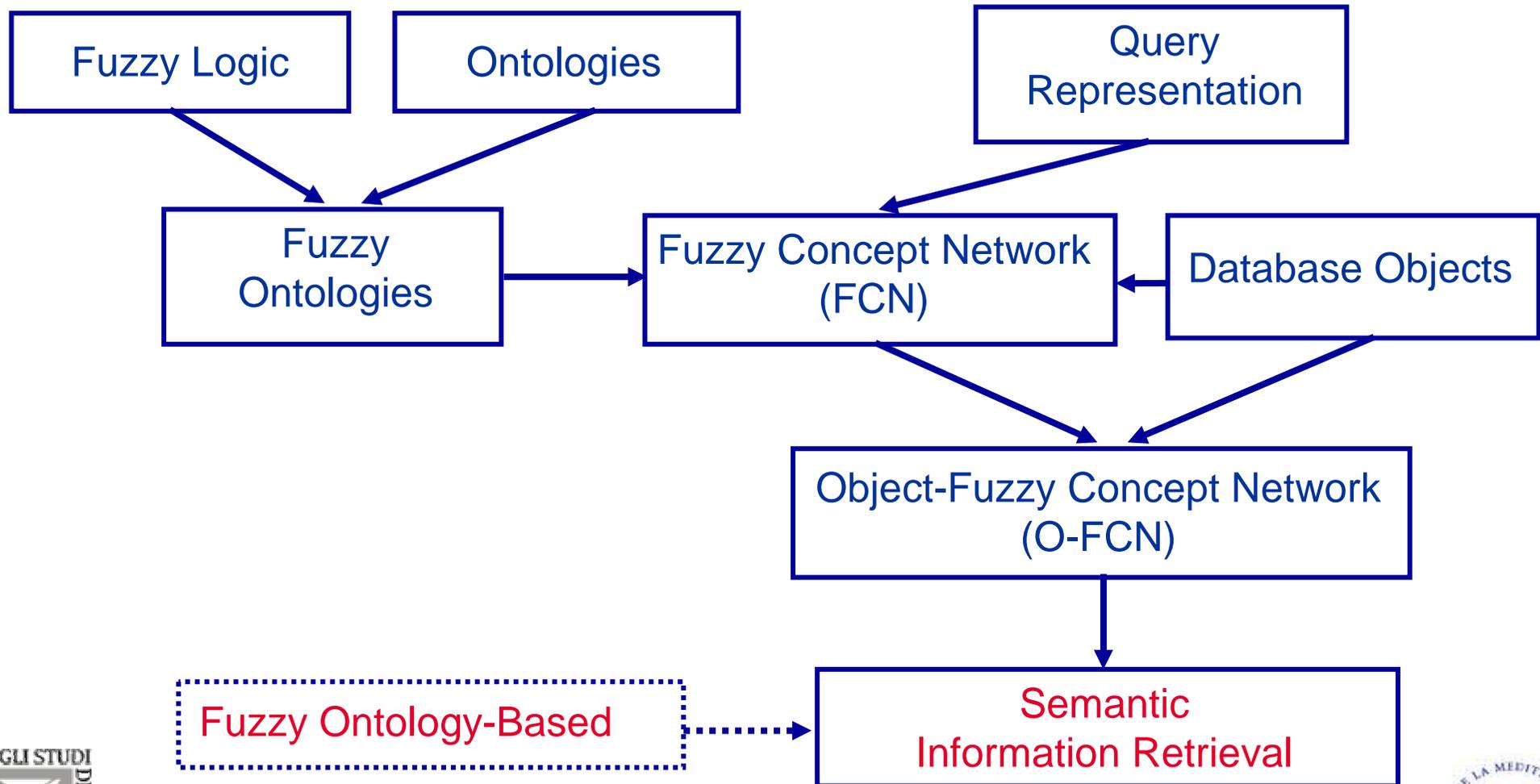
Evaluation

- *Tested within the ATELIER project (Architecture and Technologies for Inspirational Learning Environments: EU funded project)*

- *The aim was to build a digitally enhanced environment, supporting a creative learning process in architecture and interaction design education*

- *The evolution of the O-FCN mainly comes from the words in the documents inserted in a hyper-media data base and from the entities written during definition of a query by students (485 documents and 200 queries)*

- *For each query, the user had the opportunity to include up to 5 different entities and the possibility to semantically enrich his/her requests using concept modifiers: little, enough, moderately, quite, very, totally*

Fuzzy Precision

(use of 'fuzzy cardinalities')

- Precision (*Crisp*) P = Prop(R_L/R_T) = $\dfrac{|R_T \cap R_L|}{|R_T|}$

Proportion of ReTrieved documents that are ReLevant

$$\text{Precision (\textit{Fuzzy}) } P_F = \frac{\sum_{d \in Q_\theta} \mu_Q(d)}{|Q_\theta|}$$

- *Q is the fuzzy set on documents (d's), obtained as the graded result of a query*

- **Note that documents are considered as ReTrieved, above a threshold θ :** $Q_\theta = \{\, d \in D \,/\, \mu_Q(d) \geq \theta \,\} = R_T(\theta)$ *(non fuzzy set)*

Fuzzy Recall

- Recall (*Crisp*) P = Prop(R_T/R_L) = $\dfrac{|R_T \cap R_L|}{|R_L|}$

  Proportion of ReLevant documents that are ReTrieved

  $$\text{Recall } (\textit{Fuzzy}) \ R_F = \frac{\sum_{d \in Q_\theta} \mu_Q(d)}{\sum_{d \in D} \mu_Q(d)}$$

- *Relevance is a fuzzy concept:*

  All documents are ReLevant, but with a certain degree, ranging from 1 to 0

Results of evaluation

- Fuzzy Precision and Fuzzy Recall have been used in two different situations: crisp and fuzzy cases. Although results appeared to be rather similar, in the fuzzy case it has been observed a better accuracy of relevant documents

- This result was derived from the analysis of *coefficient variance*, based on fuzzy precision measure:

$$CV_P = (\sigma / P_F) \times 100$$

where σ is the Standard Deviation calculated on the relevance of the documents and P_F is Fuzzy Precision

- The fuzzy case approach identified more refinement and accuracy than the crisp case
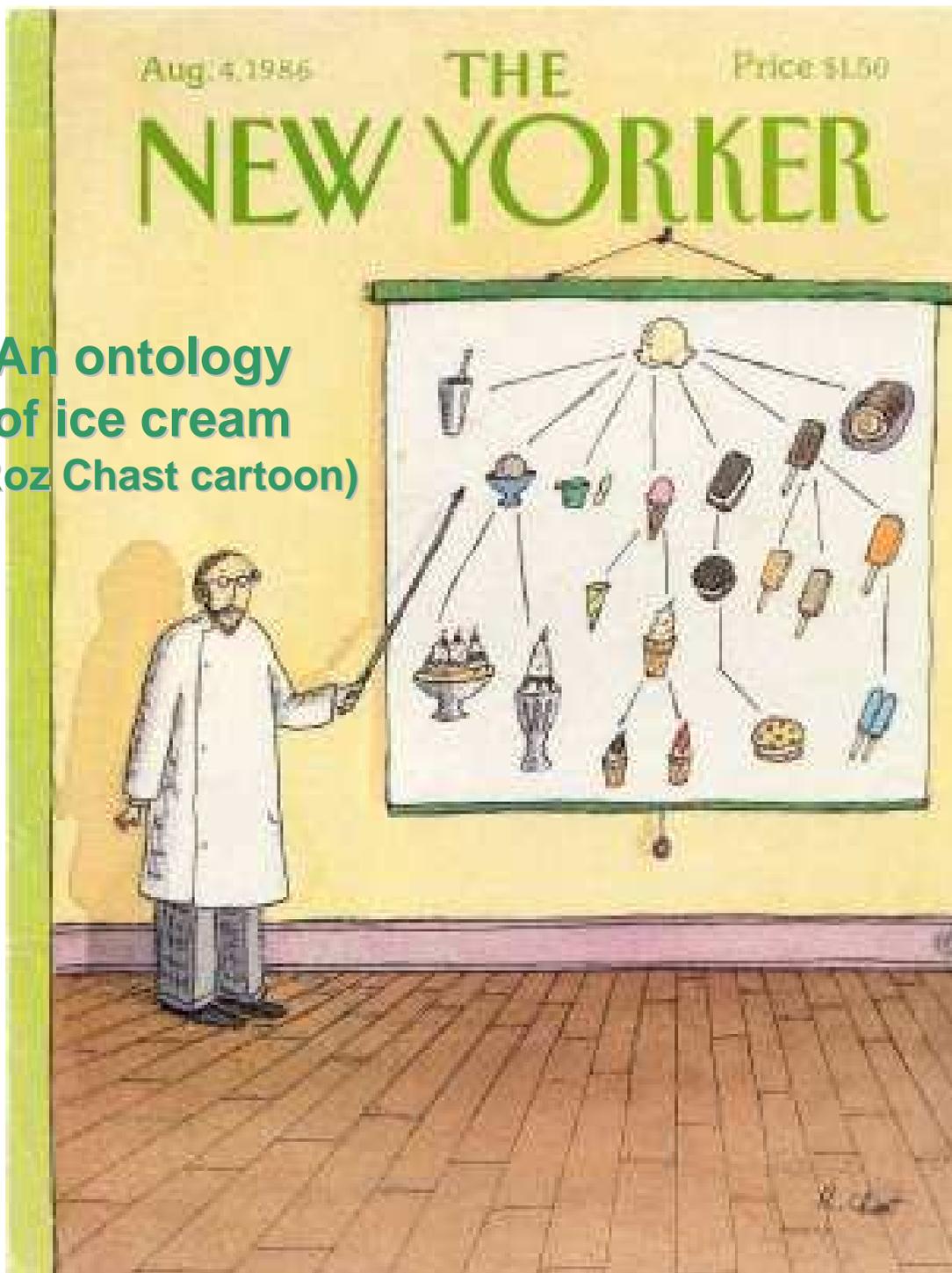
Fuzzy Ontologies in Web Information Retrieval

- *Main challenges and research directions*

  - ➤ *Incorporation of information in documents and queries into the ontologies*

  - ➤ *Improvement of retrieval through the use of relations between entities in the ontologies*

  - ➤ *Incorporation of fuzzy relationships into ontologies*

  - ➤ *Retrieval of documents from the use of the full potential of entities semantically associated in relations, instead of from only the occurrence of terms in queries*

An ontology
of ice cream
(a Roz Chast cartoon)

Thank you
for your
Attention !