



No (e-)Democracy without (e-)Knowledge

Giovanni M. Sacco

Dipartimento di Informatica, Università di Torino

Corso Svizzera 185, 10149 Torino, Italy
sacco@di.unito.it

Where is the knowledge we lost in information?

T.S. Eliot, The Rock



IGNORANTIA LEGIS NON EXCUSAT

ignorance of the law does not excuse

But how can citizens find their way in an ever-growing maze of

laws, regulations and opportunities

arising at

local, regional, national and transnational levels?



No DEMOCRACY WITHOUT KNOWLEDGE

How can citizens actively participate in law-making and decision-making, and make their voices heard, if they are not aware of government actions, proposals and decision that affect them?




No DEMOCRACY WITHOUT KNOWLEDGE

A growing number of EU citizens are becoming disillusioned with politics, which is perceived as far away and not really concerned with their problems.

Part of the problem is a **communication gap** between governments and citizens.

How can governments inform citizens of their actions?



The electronic availability of documents is a great opportunity
for democracy

But only as long as

**citizens are able to find what they need and be informed of
new opportunities, proposals, regulations...**



**ONE OF THE MOST IMPORTANT OPPORTUNITIES AND
REQUIREMENTS OF E-GOVERNMENT IS**

TO MAKE INFORMATION AVAILABLE AND FIND-ABLE

WHAT INFORMATION:

Laws, Regulations, Opportunities (Funding, ...), Strategies, etc.

Critical for democracy

BUT ALSO

Job Placement

Local Yellow Pages for trade promotion

Tourist information (hotels, restaurants, ...)

How to guides for citizens

Important for everyday life




HOW:

PULL: user interacts with the information base to find current information

PUSH: system gathers user interests and emails/sms'es them when new material arrives



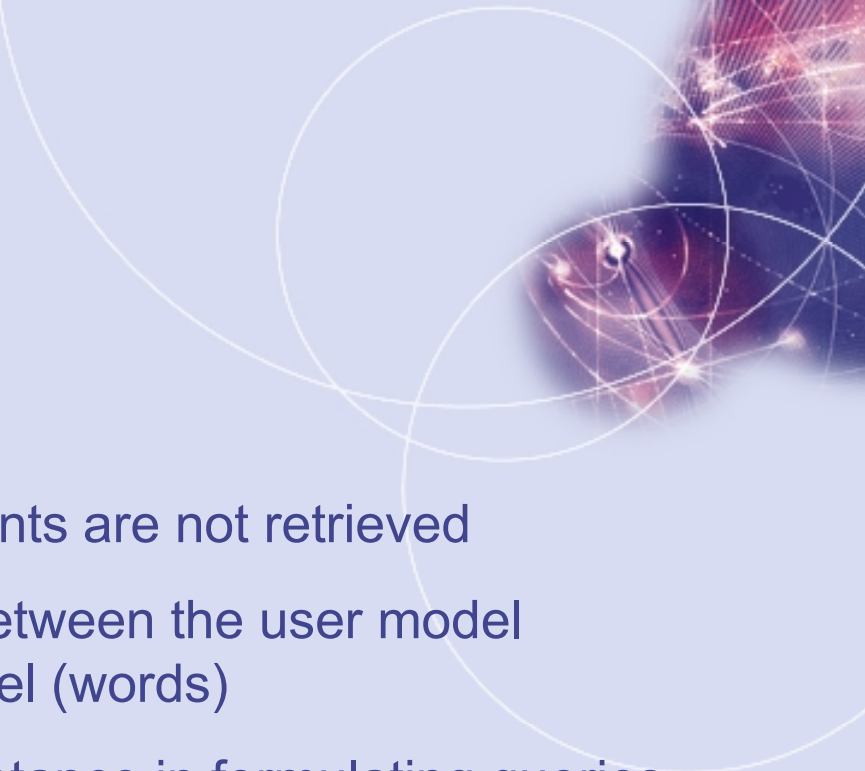
TRADITIONAL SEARCH TECHNIQUES DO NOT WORK



Since the vast majority of normative material is essentially textual and unstructured in nature

information retrieval techniques are extensively used both in pull and push strategies

BUT...

- 
1. almost 80% of relevant documents are not retrieved
 2. extremely wide semantic gap between the user model (concepts) and the system model (words)
 3. users have no or very little assistance in formulating queries
 4. results are presented as a flat list with no systematic organization: browsing is difficult or impossible.



RICH SEMANTIC SCHEMATA

1. End-users do not understand them
2. Agent mediators required: costly to implement, not transparent, hard to understand what they do
3. Schemata hard to design and maintain

PROBLEM

Traditional research has focussed on
RETRIEVAL OF INFORMATION

BUT

The most common task is BROWSING:

FIND RELATIONSHIPS

THIN ALTERNATIVES OUT



Finding a funding opportunity

Finding the laws and regulations that apply

Finding a job

BUT ALSO

Buying a digital camera

Finding a restaurant for tonight

Finding the cause of a malfunction

Selecting a photo

Finding a suspect/missing person from a photobank

....

require that the user:

1. Finds all the possible features, e.g. *Sectors of Activity, Geographical Locations, Beneficiaries, Issuers* etc.
2. Weighs all these features and freely focus on the most relevant one (e.g. *Geographical Location="Northern Italy"*)
3. Explores and finds all related features (e.g. *specific sectors, beneficiaries, issuers, etc. for Northern Italy*)
4. Repeats the process until the number of selected items is sufficiently small for manual inspection

IN SHORT

A DIFFERENT INFORMATION ACCESS PARADIGM

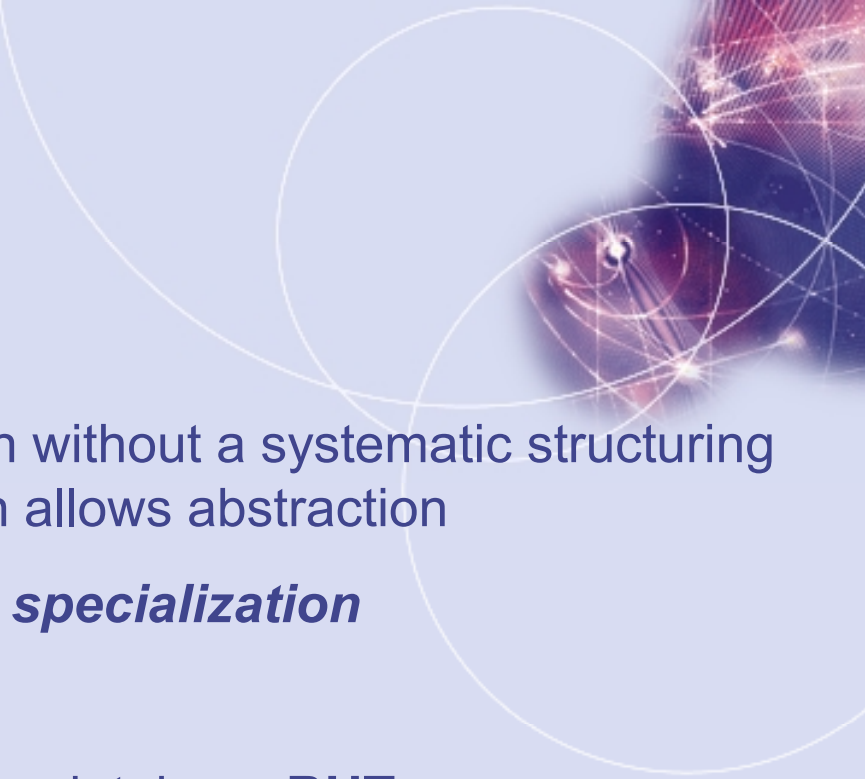
**GUIDED EXPLORATION
AND
INFORMATION THINNING**

GOALS

1. PULL: **guided exploration** of complex (normative) infobases must support conceptual access and easy and transparent end-user interactive access
2. PUSH: gather user interests and proactively inform them needs to be precise and focused otherwise perceived as junk mail



BROWSING AND GUIDED SEARCHES



There is no knowledge nor exploration without a systematic structuring
of information which allows abstraction

generalization and specialization

We need a **taxonomic** structure for our database **BUT...**



... traditional taxonomies suffer from two major (and related) problems:

1. They do not preserve relationships among concepts.
2. They do not scale up well for larger databases

SOLUTION:

DYNAMIC TAXONOMIES

Sacco, G.M., “Dynamic taxonomies: a model for large information bases”, *IEEE Trans. on Data and Knowledge Engineering*, May/June 2000

US Patent n. 6,763,349 (EU pending)

Representation

Intension: The infobase is described by a taxonomy designed by an expert (the schema)

Extension: Documents can be classified at any level of abstraction and each document is classified under n concepts ($n > 1$)

No relationships other than subsumptions (IS-A, PART-OF) need to be represented in the schema.

What is a concept?

A concept is a label which identifies a set of documents (classified under that concept)

A nominalistic approach: concepts are described by instances rather than by properties

Subsumptions require that an inclusion constraint is maintained:

If $D(C)$ denotes the set of documents classified under C and C' is a descendant of C in the hierarchy, $D(C') \subseteq D(C)$

How do concepts relate?

By IS-A

By the **Extensional Inference Rule**:

Two concepts C and C' are related if there is at least a document D which is classified both under C and C' or one of their descendants

Because of the inclusion constraint, IS-A relationships are a special case of the Extensional Inference Rule.

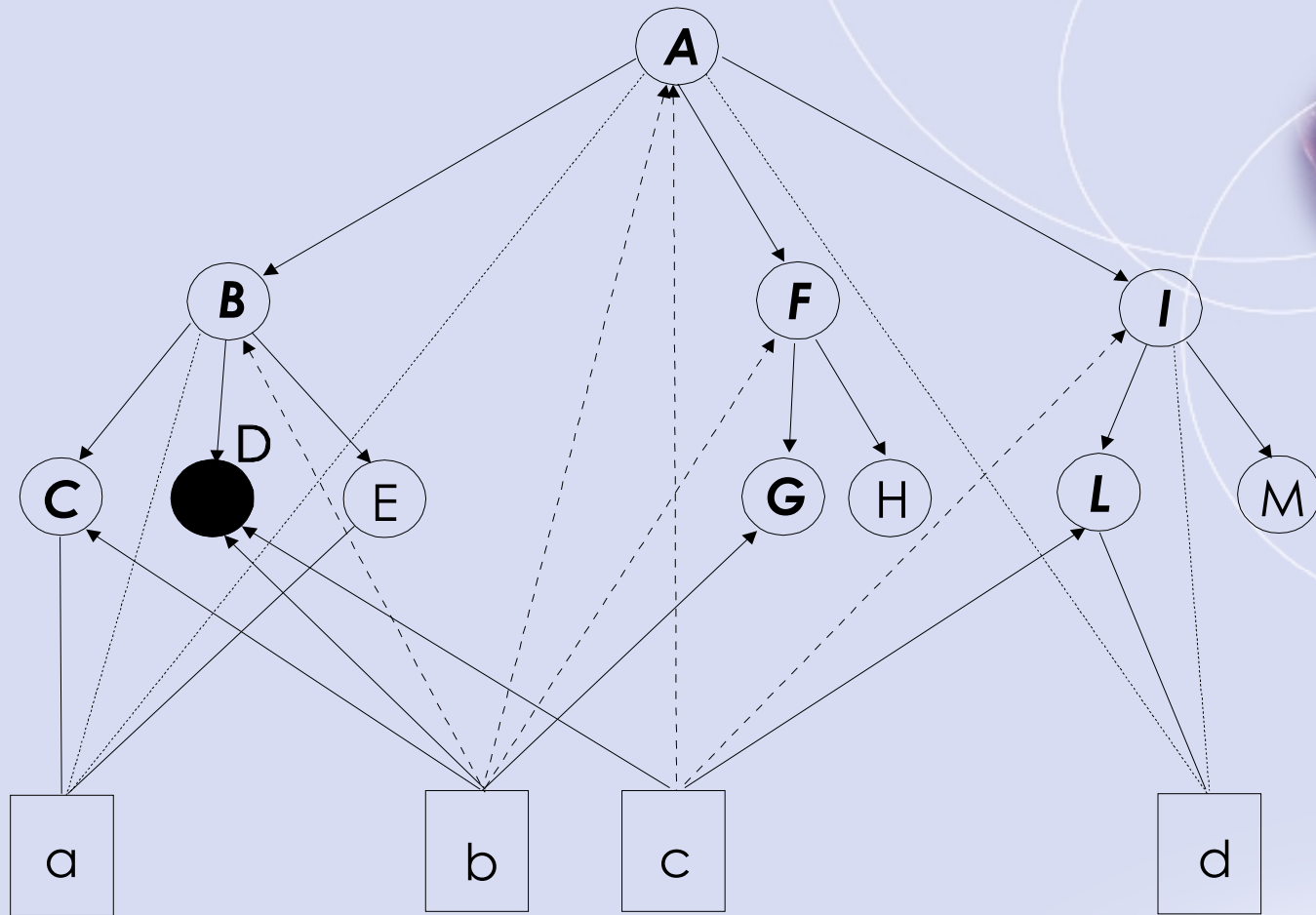


Figure 6 - Extensional inference of all concepts related to D

C, G, and L (and their ancestors) are all related to D

Important consequence:

Relationships among concepts need not be anticipated but can be inferred from the actual classification

Advantages:

a simpler schema

adapts to new relationships (dynamic)

finds unexpected relationships (discovery)

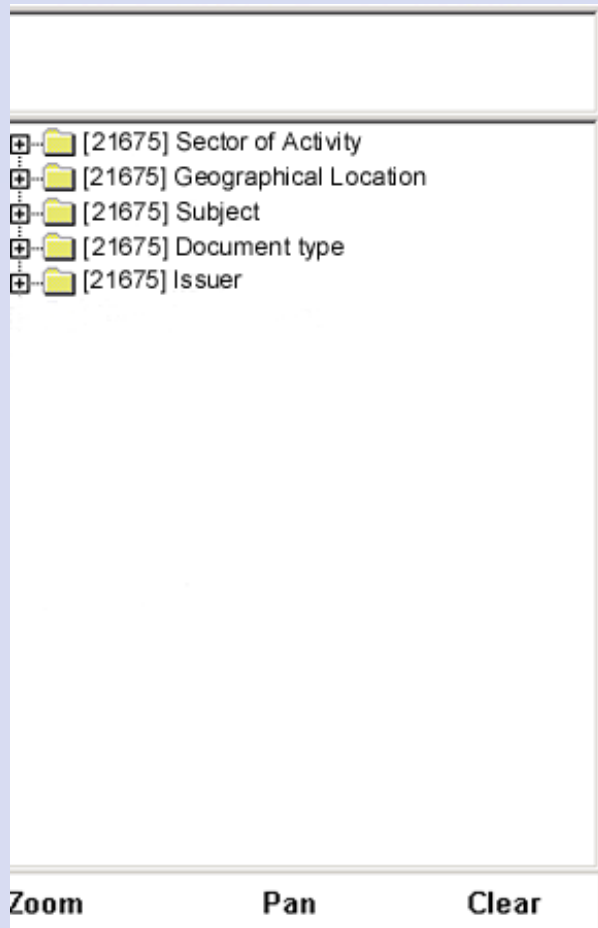


Putting it all together...

The browsing/retrieval system

1. Initial step

Tree picture of the entire infobase

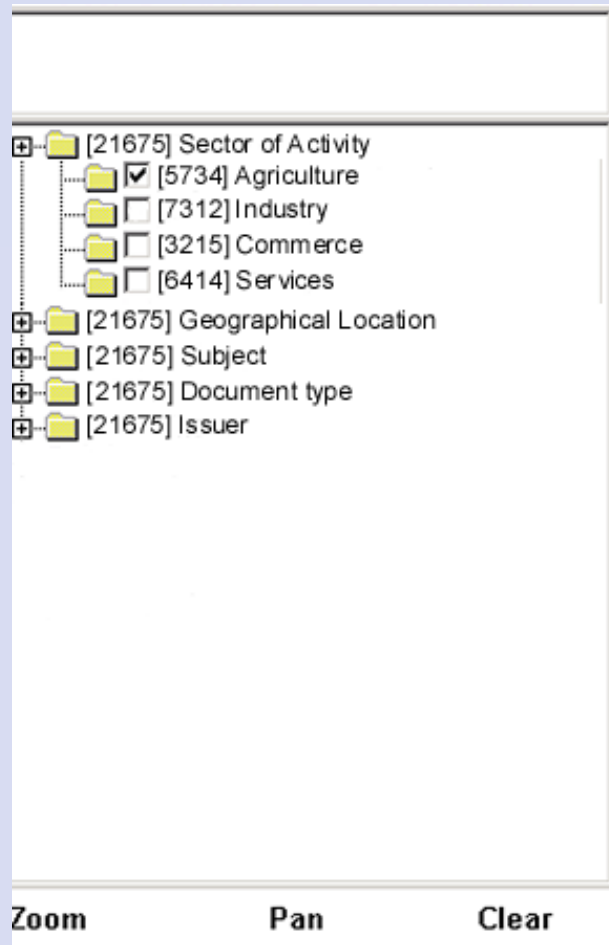


The infobase schema is used for browsing

The initial focus is the entire infobase

All information (laws, regulations, opportunities, etc.) about all European sectors

2. Select a focus (one or more concepts)

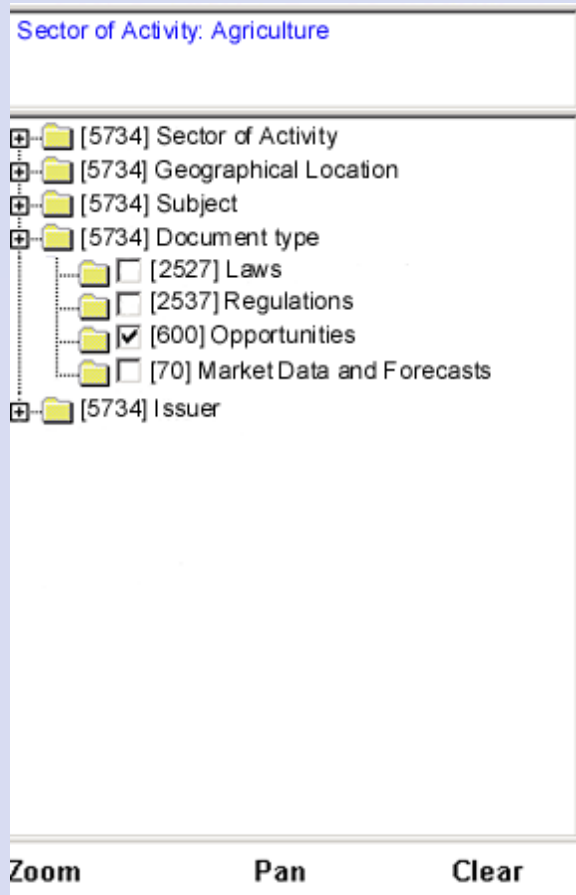


If more than one concept are selected, the document set is the union of the sets of the selected concepts.

(all set operations are supported)

Focus on Agriculture

3. Zoom and see related concepts



This is the central operation:

1. The new focus is ANDed with the previous focus
2. The entire infobase is reduced to the documents in the current focus
3. The taxonomy is reduced in order to show all and only those concepts which are extensionally related to the selected focus (filtering)


Here one of the concepts related to the new focus (Document Type) is expanded, and we are preparing to zoom on Opportunities

4. Iterate until the number of documents is sufficiently small

Sector of Activity: Agriculture AND
Document Type: Opportunity

- [600] Sector of Activity
 - [600] Geographical Location
 - [558] European Union
 - [102] Northern Europe
 - [207] Central Europe
 - [271] Southern Europe
 - [57] Greece
 - [64] Italy
 - [100] Portugal
 - [53] Spain
 - [54] Islands
 - [42] Extra EU
 - [600] Subject
 - [600] Document type
 - [600] Issuer

Zoom Pan Clear



3 zoom operations select an average 10 documents on infobases with 1,000,000 documents, described by a compact taxonomy with 1,000 concepts.

BENEFITS

- Simple and familiar interface (the only new operation is the Zoom, which is easily understood)
- The user is effectively guided to reach his goal: at each stage he has a complete list of all related concepts (i.e. a complete taxonomic summary of his current focus)
- Completely symmetric interaction: if A and B are related, the user will find B if he zooms on A, and A if he zooms on B (most systems are asymmetric)
- Discovery of unexpected relationships

BENEFITS

- Easy multilingual support (just translate concept labels)
- Easy multimedia support (documents are abstract entities: text, video, sound, whatever)
- Simple and minimal schemata:
Subsumptions only, no need for compound subjects: icons in Latvia needs not be a concept because it can be derived from the intersection of icons and Latvia

BENEFITS

- Simple integration with other retrieval techniques (IR, DB):


dynamic taxonomies as a prefilter:
they establish the context for the query

dynamic taxonomies as a conceptual summary:
they summarize long result lists



PERSONALIZATION AND PUSH

No (e-)Democracy without (e-)Knowledge



The user interest is defined by a boolean expression E on concepts, e.g.

Sector of Activity > Agriculture > Corn AND
Geographic Location > Southern Europe > Italy > Piedmont >
Cuneo



PERSONALIZATION:

E defines the initial focus instead of the universe, so that the user will not see irrelevant documents

PUSH:

new documents satisfying E will be sent to the user

CURRENT RESEARCH

Types of inheritance

Backward inheritance: if d is classified under C , it is also classified under all of C 's ancestors [ALWAYS]

Forward inheritance: if d is classified under C , it is also classified under all of C 's descendants [OFTEN]

CURRENT RESEARCH

User-friendly front-end to rich semantic infobases

Now: automatic conversion of relational database views to DT

Work in progress: automatic conversion of rich semantic schemata to DT