

# HCSNet

Next Generation Search Technology

Shlomo Geva

QUT

# Translating Natural Language Queries for use in XML Information Retrieval Systems

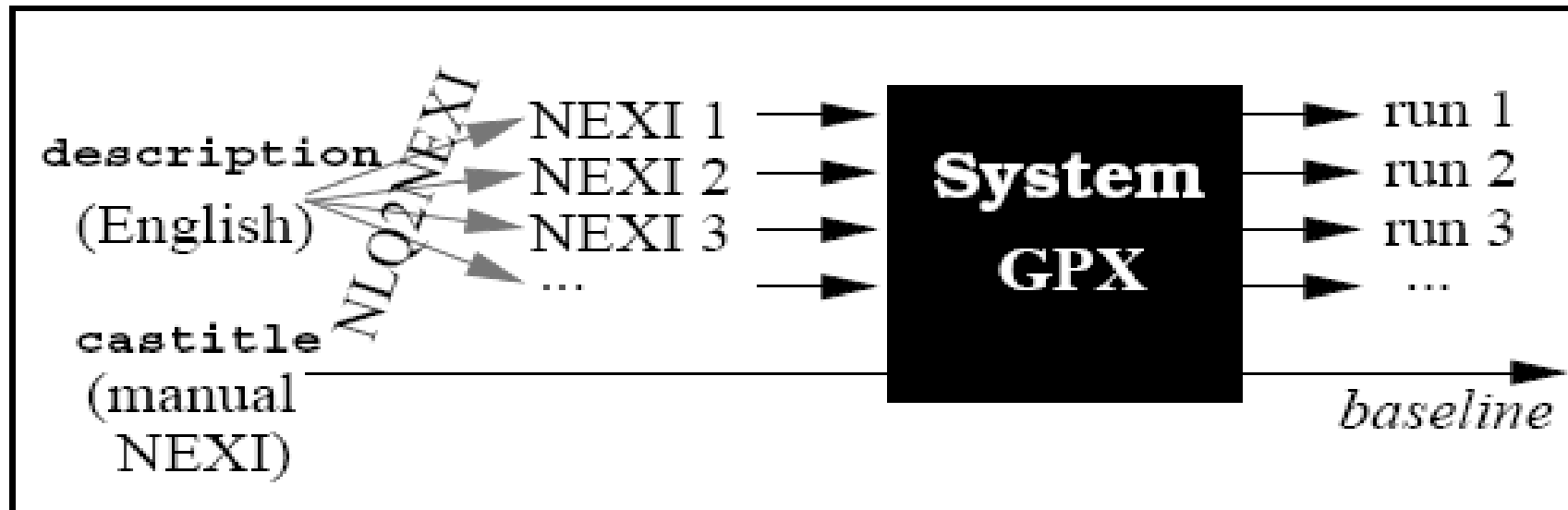
What we tell the search engine

```
<castitle>  
  //article[about(./p,"data embedding")]  
  //p[about(.,watermarking)]  
</castitle>
```

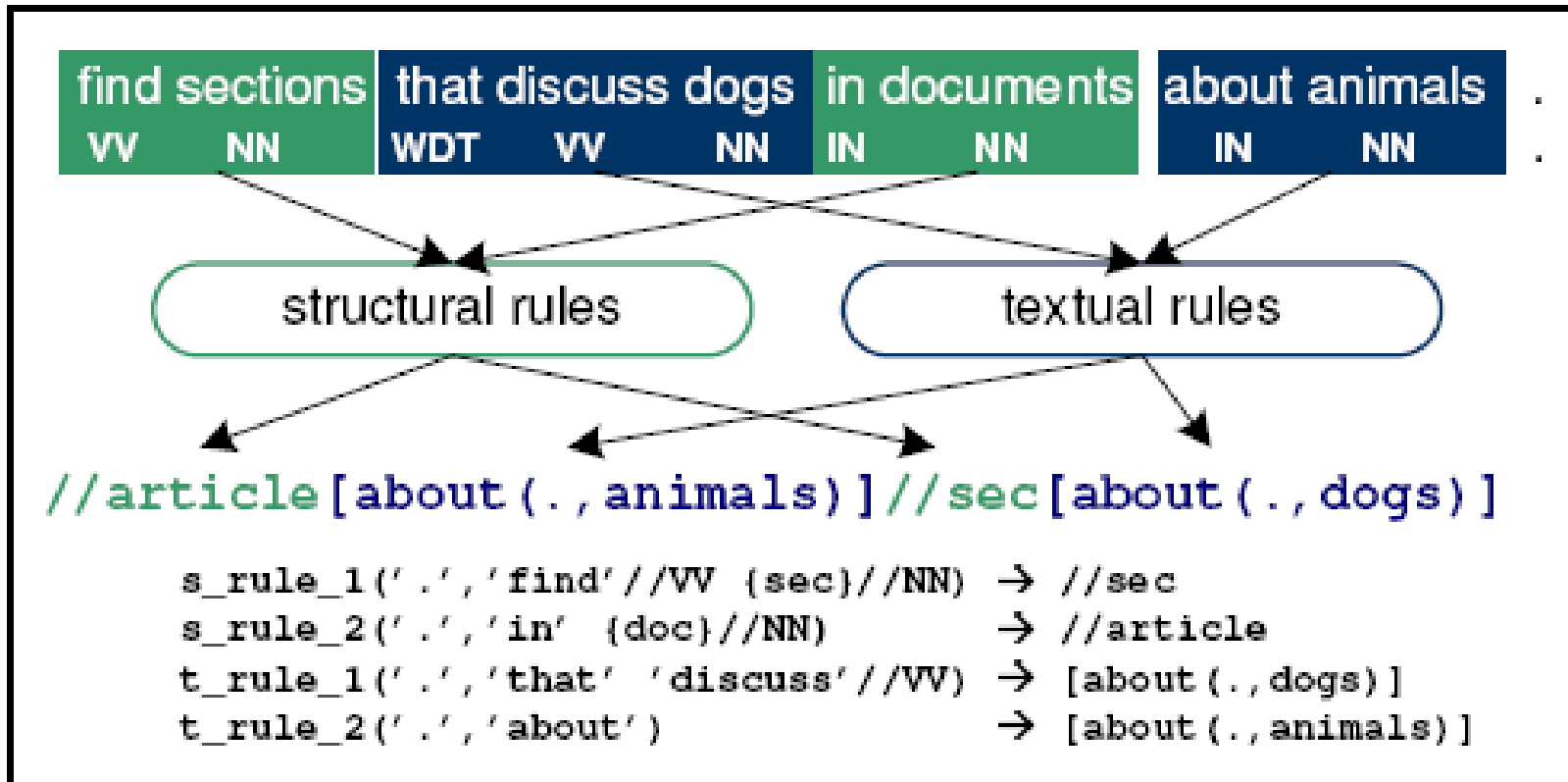
What the user really wanted to say

```
<description>  
  We are looking for paragraphs describing  
  watermarking in articles which describe  
  data embedding.  
</description>
```

# NLQ2NEXI



# Rule Templates

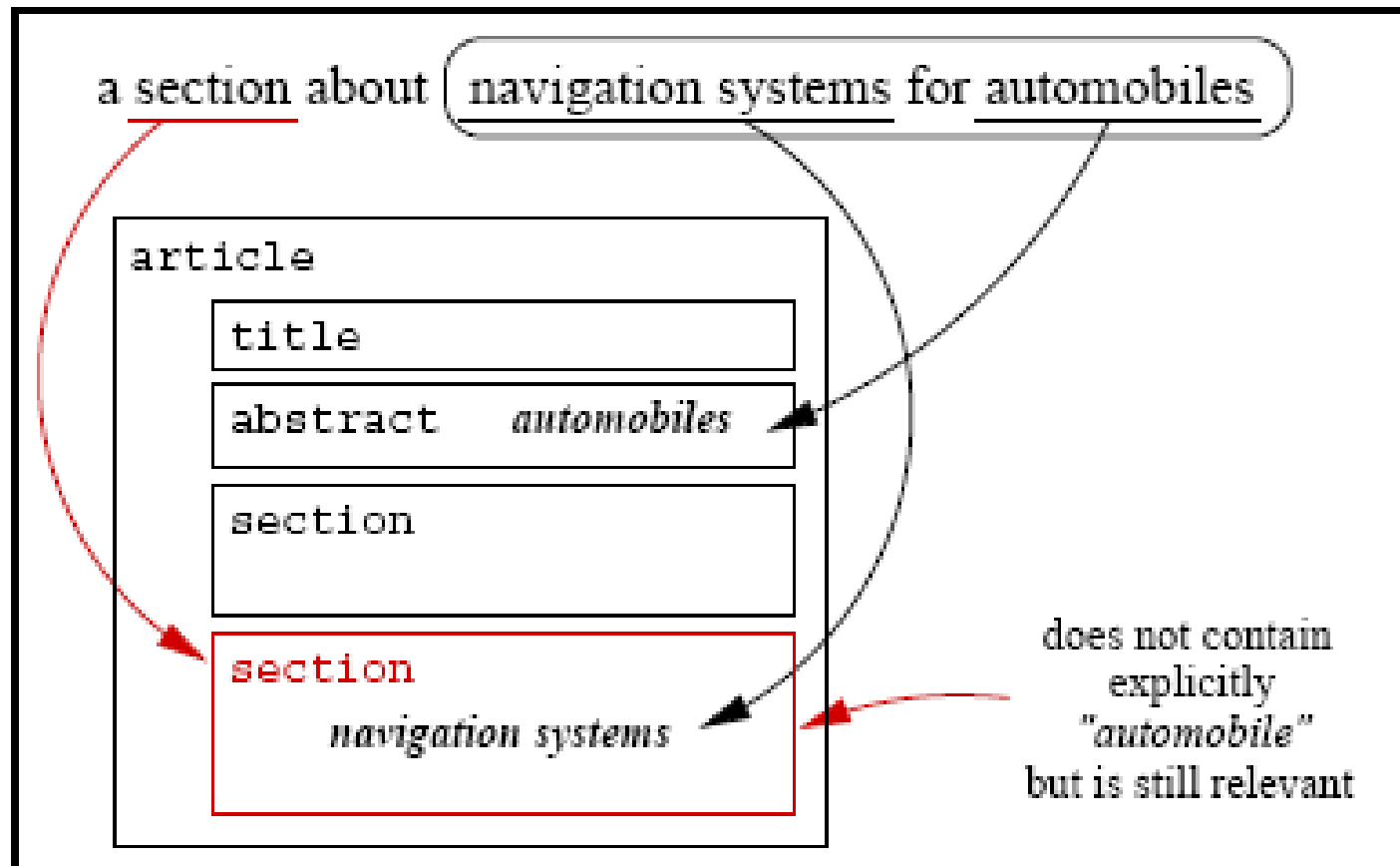


# Lingusitic constructions

to recognize some frequent linguistic constructions that implicitly refer to structure

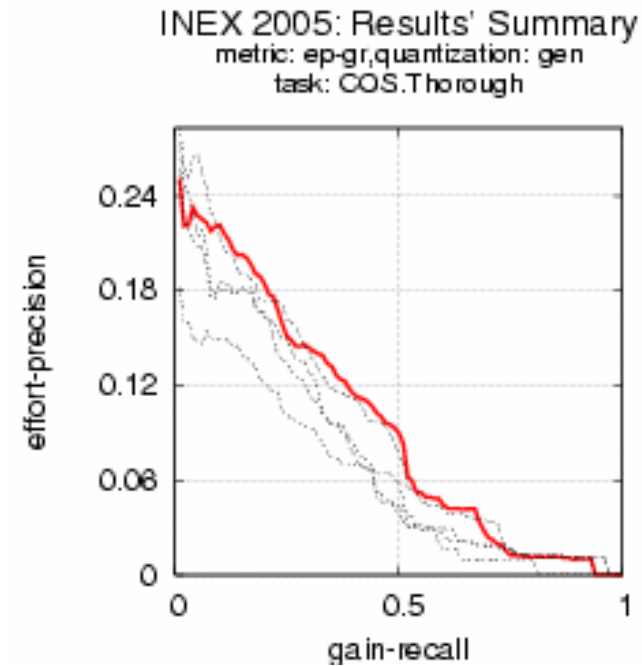
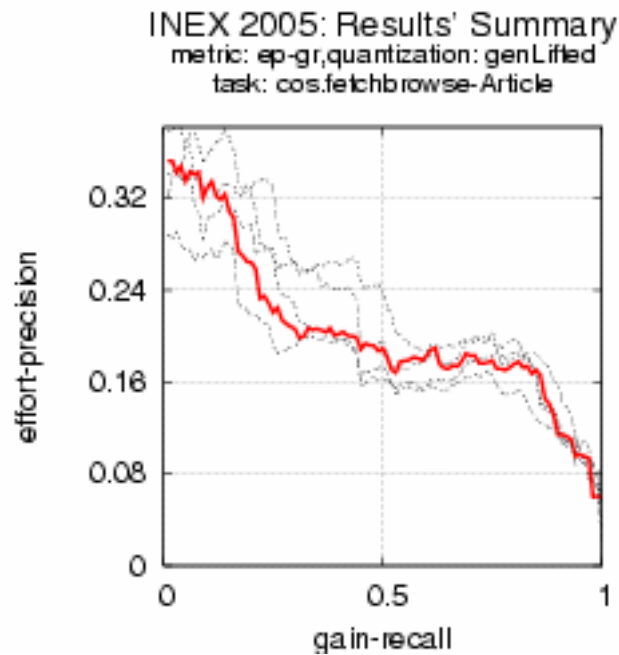
- (the agent of the verb “to write” is an author (<*au*>)
- the object of “to cite” is a bibliographic item (<*bb*>)
- Unlike database querying, XML NLI do not require domain-dependant knowledge (although this is a potentially fruitful path to follow).

# Contextual search



# Results from INEX 2005

Overall the NLQ2NEXI (description based) was approximately 80% as effective as the baseline (formal NEXI) as measured by MAP value.



# Collaborative P2P IR

(beyond federated desktops)

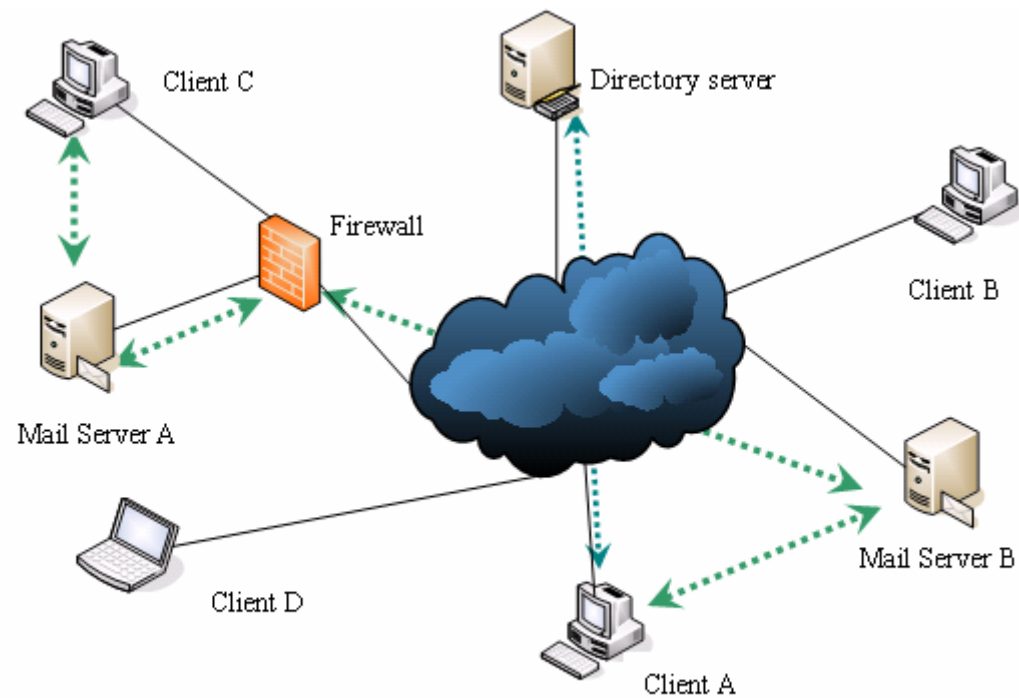


Figure 1. System overview

# Collaborative Collections

- Search and curate distributed collaborative collections
- Extensive system support (mark-up, suggest, inform, alert, ...)
- Advanced search facilities, including Text, XML and Multimedia
- Ability to tunnel through firewalls hence it must be certifiably secure.
- Store and forward modus operandi = reliability & persistence
- Complete access control by end-users = security & privacy (documents are not necessarily in the public domain!)
- Emergent group / collection formation through adaptive learning to combat network flooding