An Improved Semantic Search Model Based on Hybrid Fuzzy Description Logic

Ruixuan Li, Kunmei Wen, Zhengding Lu, Xiaolin Sun, Zhigang Wang

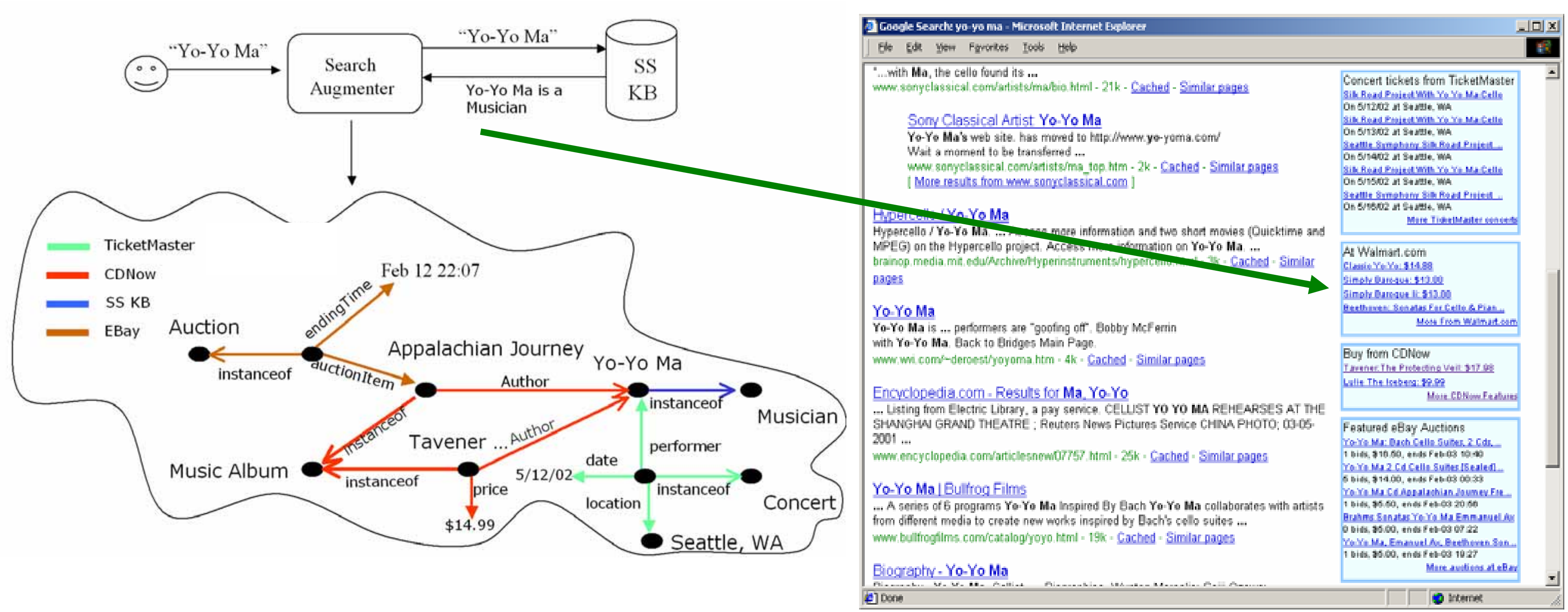
Huazhong University of Science and Technology

rxli@hust.edu.cn

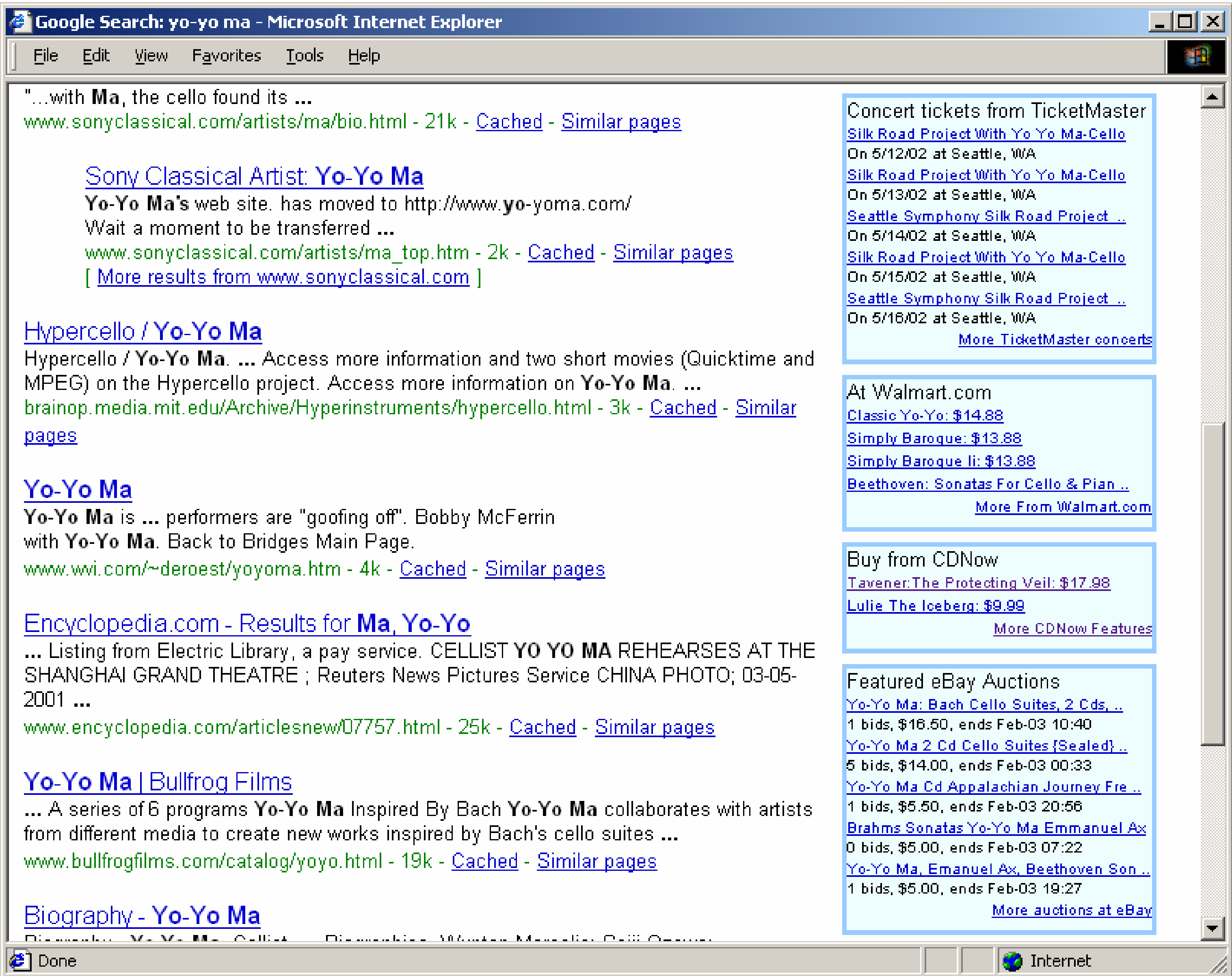
Background

- **Semantic search (R. Guha, WWW 2003)**
- **Semantic Web + Search Engine**
  - **finds out the internal knowledge in Knowledge Base (KB) to improve the search results gained by current search engines**
  - **evolves to next generation of search engines built on Semantic Web**
- **Two types of search**
  - **searching the current Internet**
  - **searching the Semantic Web portals**

Related Work-Tap



Tap


Google Search: yo-yo ma - Microsoft Internet Explorer
File Edit View Favorites Tools Help
"...with Ma, the cello found its ...
www.sonyclassical.com/artists/ma/bio.html - 21k - Cached - Similar pages
Sony Classical Artist: Yo-Yo Ma
Yo-Yo Ma's web site. has moved to http://www.yo-yoma.com/
Wait a moment to be transferred ...
www.sonyclassical.com/artists/ma_top.htm - 2k - Cached - Similar pages
[More results from www.sonyclassical.com]
Hypercello / Yo-Yo Ma
Hypercello / Yo-Yo Ma. ... Access more information and two short movies (Quicktime and MPEG) on the Hypercello project. Access more information on Yo-Yo Ma. ...
brainop.media.mit.edu/Archive/Hyperinstruments/hypercello.html - 3k - Cached - Similar pages
Yo-Yo Ma
Yo-Yo Ma is ... performers are "goofing off". Bobby McFerrin
with Yo-Yo Ma. Back to Bridges Main Page.
www.wwi.com/~deroest/yoyoma.htm - 4k - Cached - Similar pages
Encyclopedia.com - Results for Ma, Yo-Yo
... Listing from Electric Library, a pay service. CELLIST YO YO MA REHEARSES AT THE SHANGHAI GRAND THEATRE ; Reuters News Pictures Service CHINA PHOTO; 03-05-2001 ...
www.encyclopedia.com/articlesnew/07757.html - 25k - Cached - Similar pages
Yo-Yo Ma | Bullfrog Films
... A series of 6 programs Yo-Yo Ma Inspired By Bach Yo-Yo Ma collaborates with artists from different media to create new works inspired by Bach's cello suites ...
www.bullfrogfilms.com/catalog/yoyo.html - 19k - Cached - Similar pages
Biography - Yo-Yo Ma
Concert tickets from TicketMaster
Silk Road Project With Yo Yo Ma-Cello
On 5/12/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/13/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/14/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/15/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/16/02 at Seattle, WA
More TicketMaster concerts
At Walmart.com
Classic Yo-Yo: $14.88
Simply Baroque: $13.88
Simply Baroque Ii: $13.88
Beethoven: Sonatas For Cello & Pian ..
More From Walmart.com
Buy from CDNow
Tavener:The Protecting Veil: $17.98
Lulie The Iceberg: $9.99
More CDNow Features
Featured eBay Auctions
Yo-Yo Ma: Bach Cello Suites, 2 Cds, ..
1 bids, $16.50, ends Feb-03 10:40
Yo-Yo Ma 2 Cd Cello Suites {Sealed} ..
5 bids, $14.00, ends Feb-03 00:33
Yo-Yo Ma Cd Appalachian Journey Fre ..
1 bids, $5.50, ends Feb-03 20:56
Brahms Sonatas Yo-Yo Ma Emmanuel Ax
0 bids, $5.00, ends Feb-03 07:22
Yo-Yo Ma, Emanuel Ax, Beethoven Son ..
1 bids, $5.00, ends Feb-03 19:27
More auctions at eBay
Done
Internet

Related Work

- **Swoogle is a prototype system of information retrieval (IR). The search results are physical documents on Semantic Web (such as RDF and OWL files).**
  - **It has not used the semantic structure information in documents.**
- **Turing center in the University of Washington develops the system KnowItAll to extract the information on the Web.**
  - **Its long-term aim is to replace the search engine by information extraction.**

Related Fields

- **Description Logic (DL): Tableaux algorithm & many optimized.**
- **Classical DLs can only define the certain concepts and properties, which cannot solve the fuzzy problem of ontology system. Fuzzy DL are designed to expand the classic DLs to make it more applicable to ontology system.**
- **Ranking the search results: the number of relationships between entities in a KB will be much larger than the number of entities themselves.**

Our Focus

- Combine techniques of semantic reasoning and information retrieval
- Improve semantic search model by supporting imprecise and fuzzy search

Outline

- **Semantic Search Model**
- **Architecture of Semantic Search Model**
- **RBAC Security Ontology**
- **Experiment and Evaluation**
- **Conclusion**

Basic Model

- **Idea: integrating search and inference**
  - **Formal DL reasoning: search the resources and relationships**
  - **Traditional IR: locate the exact resource**
- **Basic model:**
  - **Triple[D, Q, R(d_i, Q)]**
    - **D is the set of documents d_i**
    - **Q is a query**
    - **R(d_i, Q) is the similarity between d_i and Q, here $d_i \in D$**
  - **Here documents and queries are modeled as individuals and concepts respectively**

Basic Model

- **Basic model ⟷ formal DL (ALC constructors) + traditional IR**

Constructor	Syntax	Semantics
Top (Universe)	$\top$	Δ^I
Bottom (Nothing)	$\bot$	ϕ
Atomic Concept	A	$A^I \subseteq \Delta^I$
Atomic Role	R	$R^I \subseteq \Delta^I \times \Delta^I$
Conjunction	C $\cap$ D	$C^I \cap D^I$
Disjunction	C $\cup$ D	$C^I \cup D^I$
Negation	$\neg$ C	$\Delta^I \setminus C^I$
Value restriction	$\forall R.C$	$\{a \in \Delta^I \mid \forall b \in \Delta^I, (a,b) \in R^I \rightarrow b \in C^I\}$
Full existential quantification	$\exists R.C$	$\{a \in \Delta^I \mid \exists b \in \Delta^I, (a,b) \in R^I \wedge b \in C^I\}$

Query Form

- **Different users have different privileges for different resources**
- **A query is defined as the form**
  - $Q_i = Q_{i1} \cap Q_{i2} \cap Q_{i3}$
  - Q_{i1} **means a concept**
  - Q_{i2} **is any formal query about resources or the relationships between them**
  - Q_{i3} **is a keyword query**

Query Form

Four typical queries as follows:

- **Q_{i1} : Concept Query, form as Q_{i2} = "C" where means a concept.**
- **Q_{i2}: Relationship Query, form as Q_{i2} = "C1"&"C2" where C1 and C2 are concepts.**
- **Q_{i3}: Keyword Query, form as Q_{i3} = "D" where D means a keyword which appears in the text. In fact, Q_{i3} belongs to traditional query.**
- **$Q_{i1} \cap Q_{i3}$: Conjunctive Query, form as $Q_{i1} \cap Q_{i3}$ = ("A"or"B") $\cap$ D" where A means a user, B means a role and D means a keyword.**

Reasoning Based on Description Logic

- ***Concept Reasoning*. Given $Q_i = Q_{i1}$ *where* Q_{i1} means concept, we can get all the sub concepts and all the instances are returned.**
- ***Relationship Reasoning*. Given $Q_i = Q_{i2}$ where Q_{i2} includes two concepts, we can get the relationship between them or null if there is not any relationship.**
- ***Conjunctive Query Reasoning*. In fact it integrates inference with search by providing both formal query and keyword query. Given $Q_i = Q_{i1} \cap Q_{i3}$ where Q_{i1} means user or role and Q_{i3} is a keyword query.**

Result Ranking

- $R_i = R_{i1} + R_{i2} + R_{i3}$ **for the query** $Q_i = Q_{i1} \cap Q_{i2} \cap Q_{i3}$.
- **Given** Q_{i1}**, if d is the instance of the query** Q_{i1} **then the corresponding** $R_{i1}=1$**, otherwise the value equals to 0.**
- R_{i2} **is determined by the important value of the relationship. Not deeply discussion here, need to be further explored.**
- **Use traditional tf-idf (term frequency • inverse document frequency) method to compute the value of** R_{i3}.

Extended Model

- **Extended model ⟷ formal DL + Type-1 FALC**
- **Type-1 FALC Syntax and semantics**

$$\top^I(d) = 1$$

$$\bot^I(d) = 0$$

$$(C \cap D)^I(d) = \min\{C^I(d), D^I(d)\}$$

$$(C \cup D)^I(d) = \max\{C^I(d), D^I(d)\}$$

$$\neg C^I(d) = 1 - C^I(d)$$

$$(\forall R.C)^I(d) = \inf_{d' \in \Delta^I} \max\{1 - R^I(d, d'), C^I(d)\}$$

$$(\exists R.C)^I(d) = \sup_{d' \in \Delta^I} \min\{R^I(d, d'), C^I(d)\}$$

Extended Model

- **Classic DL such as ALC cannot deal with the imprecise description**
- **Extended model: use type-1 FALC to replace IR model. The fuzzy degree is equal to the ranking value computed by traditional IR**

Complete Model

- Exist imprecise terminological axiom (TBox) and fuzzy individual membership (ABox).
- It is necessary to maintain trust degree for both of TBox and ABox.
- Complete model ←→ Type-1 FALC + Type-1 FALC

Complete Model

- **Propose type-2 FALC based on type-2 fuzzy sets, Type-2 FALC syntax and semantics**

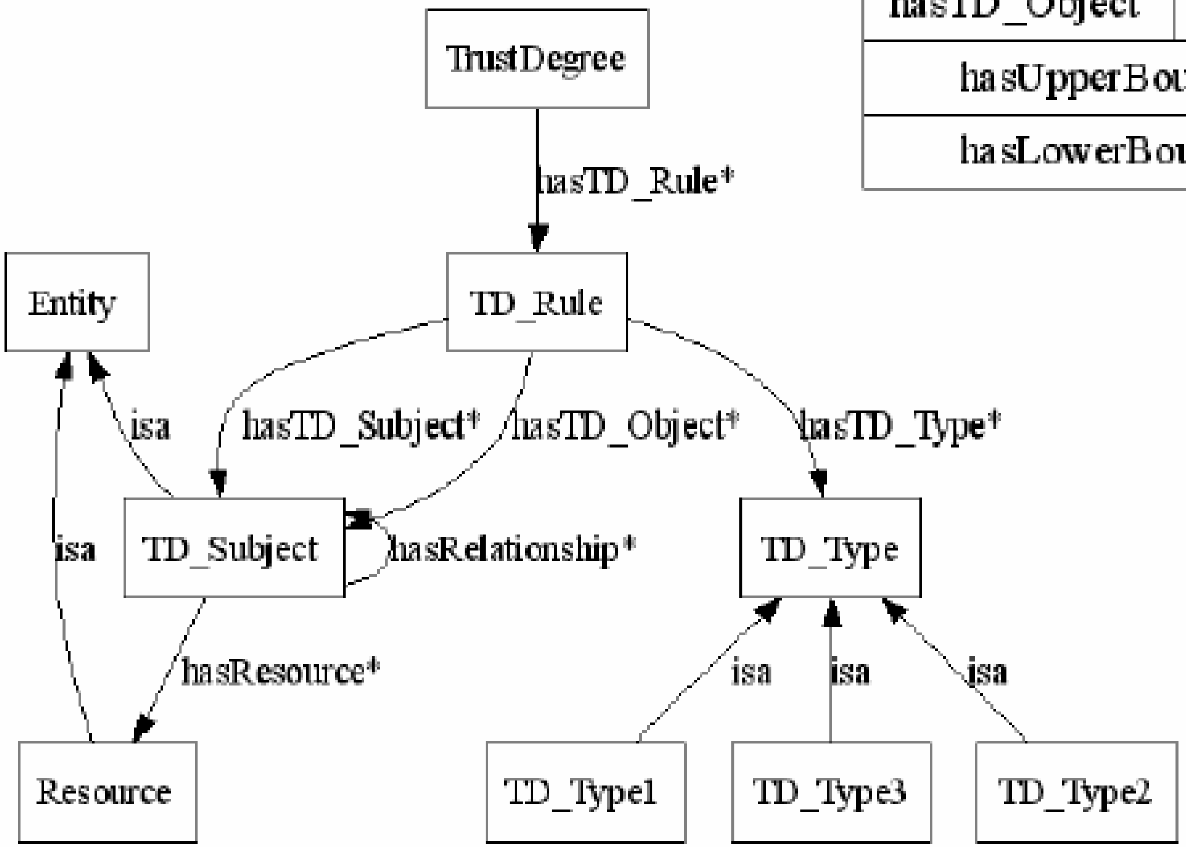
Constructor	Syntax	Semantics
Top (Universe)	$\top$	Δ^I
Bottom (Nothing)	$\perp$	ϕ
Atomic Concept	$A_{[a,b]}$	$A^I_{[a,b]} \subseteq \Delta^I$
Atomic Role	$R_{[a,b]}$	$R^I_{[a,b]} \subseteq \Delta^I \times \Delta^I$
Conjunction	$C_{[a,b]} \cap D_{[c,d]}$	$(C \cap D)^I_{[\min(a,c),\min(b,d)]}$
Disjunction	$C_{[a,b]} \cup D_{[c,d]}$	$(C \cup D)^I_{[\max(a,c),\max(b,d)]}$
Negation	$\neg C_{[a,b]}$	$C^I_{[1-b,1-a]}$
Value restriction	$\forall R_{[a,b]}.C_{[c,d]}$	$\forall y.\max(R_{[1-b,1-a]}(x,y), C_{[c,d]}(y))$
Full existential quantification	$\exists R_{[a,b]}.C_{[c,d]}$	$\exists y.\min(R_{[a,b]}(x,y), C_{[c,d]}(y))$

Complete Model

- **The complete model use type-2 fuzzy *ALC* to replace the formal DL in the extended model. In real world, there are the imprecise terminological axiom (TBox) and fuzzy individual membership (ABox)**
- **We apply type-2 fuzzy *ALC* to deal with the description in ontology for trust degree management (OntoTD)**
- **We design a simple ontology named OntoTD for trust management**

OntoTD

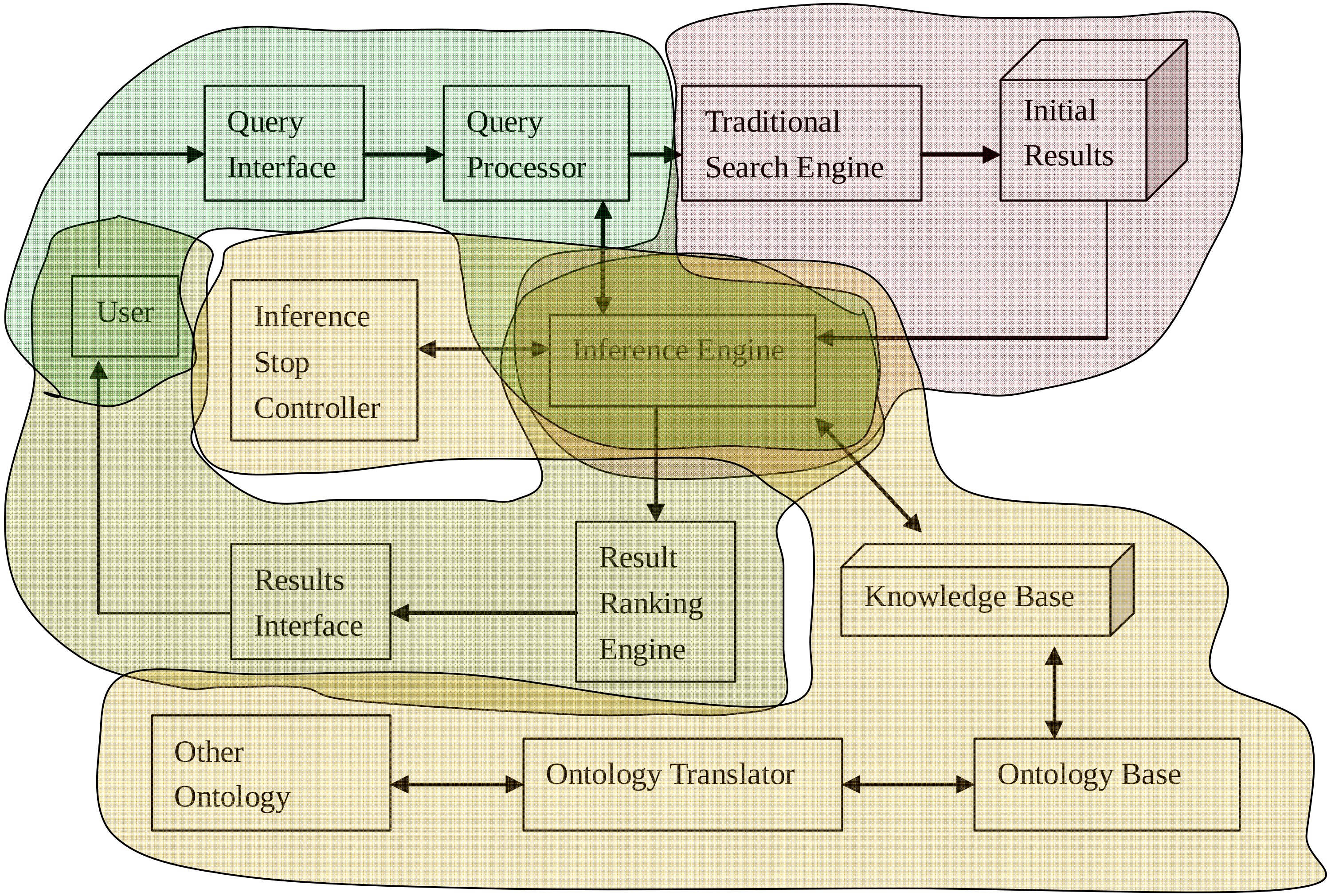
TD_Rule		
hasTD_Type	Instance*	TD_Type
hasTD_Subject	Instance*	TD_Subject
hasTD_Object	Instance*	TD_Subject
hasUpperBound	Float*	
hasLowerBound	Float*	



Outline

- **Semantic Search Model**
- **Architecture of Semantic Search Model**
- **RBAC Security Ontology**
- **Experiment and Evaluation**
- **Conclusion**

Architecture of Semantic Search Model



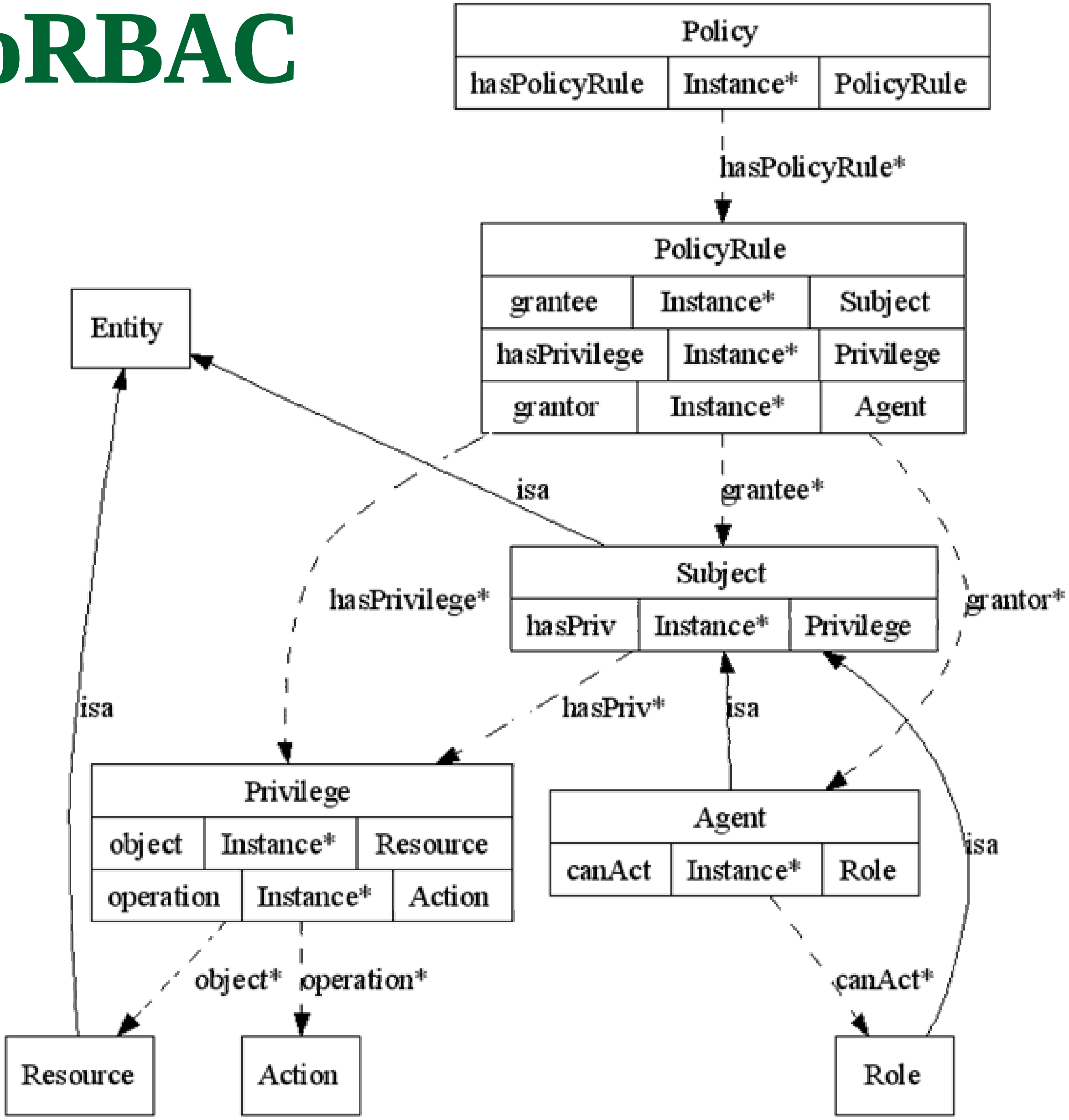
Outline

- **Semantic Search Model**
- **Architecture of Semantic Search Model**
- **RBAC Security Ontology**
- **Experiment and Evaluation**
- **Conclusion**

RBAC Security Policy

- **Great demands: Secure Search, such as Intranet search which must satisfy requirements of access control in the background of government or business.**
- **Role-Based Access Control (RBAC)**
  - **Various ways of specifying policy in the last decade**
  - **Ontology-based approach**
  - **A security ontology based on RBAC policy**

OntoRBAC



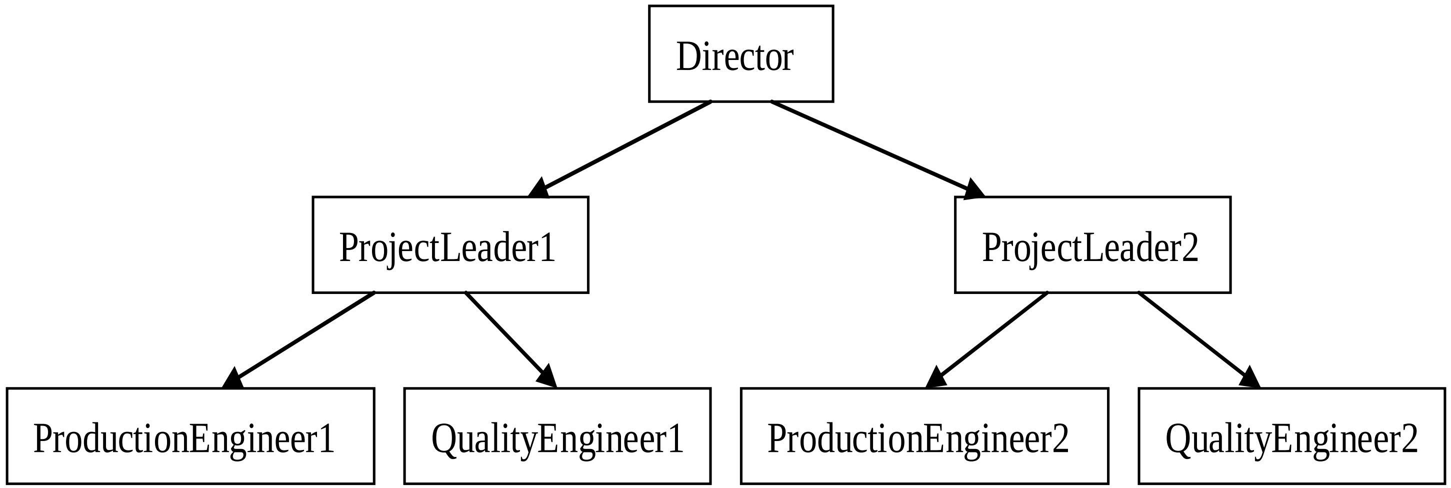
Policy
hasPolicyRule
Instance*
PolicyRule
hasPolicyRule*
PolicyRule
grantee
Instance*
Subject
hasPrivilege
Instance*
Privilege
grantor
Instance*
Agent
Entity
isa
grantee*
hasPrivilege*
Subject
hasPriv
Instance*
Privilege
grantor*
isa
hasPriv*
isa
Privilege
object
Instance*
Resource
operation
Instance*
Action
Agent
canAct
Instance*
Role
isa
object*
operation*
canAct*
Resource
Action
Role

RBAC Security Ontology

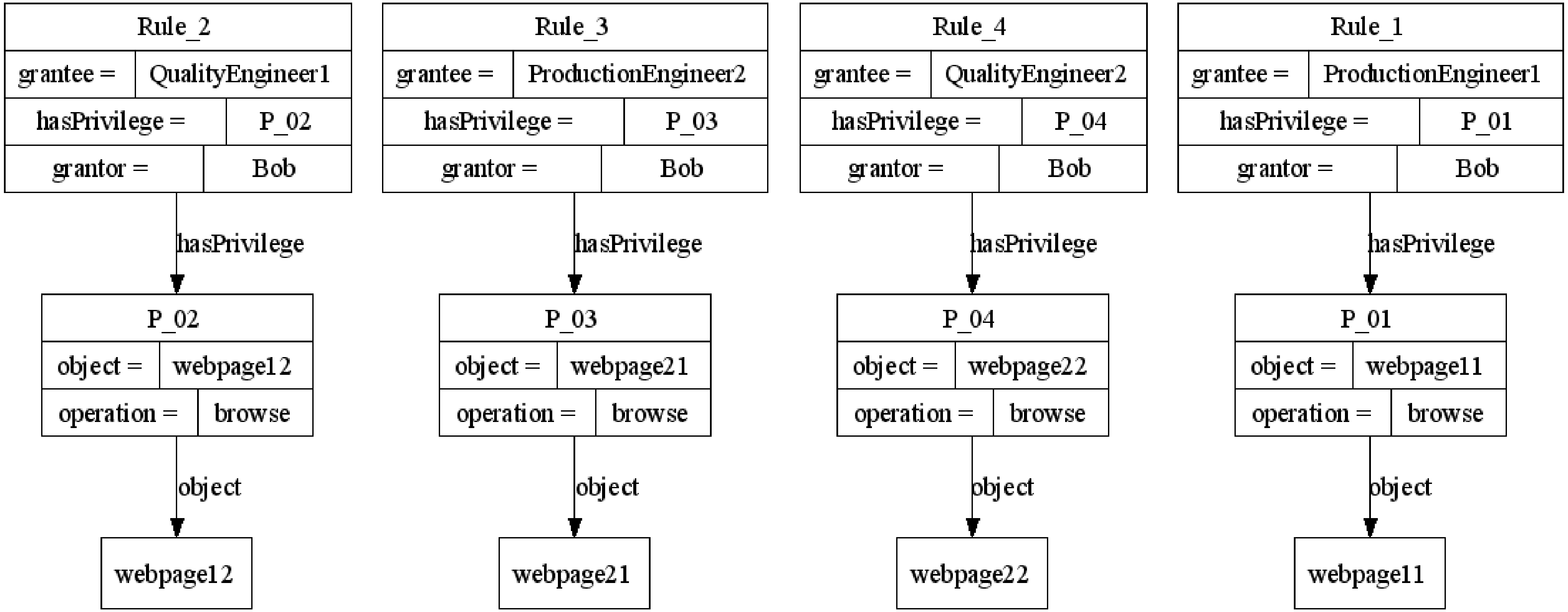
- **We implement the security ontology using OWL DL as the ontology language**

```
< owl : Classrdf : ID = ”Subject” >< rdfs : subClassOf >
< owl : Classrdf : ID = ”Entity”/ >< /rdfs : subClassOf >
< /owl : Class >< owl : Classrdf : ID = ”Role” >
< rdfs : subClassOfrdf : resource = ”#Subject”/ > ..
< /owl : ObjectP roperty >< owl : ObjectPropertyrdf : ID = ”hasP riv” >
< rdfs : rangerdf : resource = ”#Privilege”/ >
< rdfs : domainrdf : resource = ”#Subject”/ >< /owl : ObjectProperty >
```

Instance of Roles



Instance of PolicyRules



Outline

- **Semantic Search Model**
- **Architecture of Semantic Search Model**
- **RBAC Security Ontology**
- **Experiment and Evaluation**
- **Conclusion**

Experiment Settings

- **We implement Ontology Security Semantic Search Engine (Onto-SSSE) in Java**
  - **Use the Lucene as the traditional search engine based on keyword query**
  - **Use Jena as the reasoning tool based on RBAC security ontology**
  - **Only implemented basic model**

Experiment Settings

- No commonly agreed evaluation methodology and benchmark for semantic search
- The OntoRBAC evaluation dataset
  - OntoRBAC security ontology (including 12 classes, 16 properties and 20 individuals)
  - The set of campus web pages (more then 200MB)

Semantic Search Results

Query ID	Query Form	Query Item	Reasoning Type	Query Result
Q_1	Q_{i1}	"Director"	Concept Reasoning	Sub-roles:ProjectLeader1,ProductionEngineer1, QualityEngineer1,ProjectLeader2, ProductionEngineer2,QualityEngineer2; Privileges:(browse,webpage11), (browse,webpage12), ……
Q_2	Q_{i2}	"ProjectLeader1"& "ProductionEngineer1"	Relationship Reasoning	seniorRoleOf
Q_3	$Q_{i1} \cap Q_{i3}$	"Director & computer"	Conjunctive Query Reasoning	webpage list: webpage11, webpage21…. Where include the text "computer" in these web pages

Comparison with Traditional Search

Query form	Reasoning Type	Traditional method	Semantic search
Q_{i1}	Concept Reasoning	Not support	Support
Q_{i2}	Relationship Reasoning	Not support	Support
Q_{i3}	No Reasoning (keyword search)	Support	Support
$Q_{i1} \cap Q_{i3}$	Conjunctive Query Reasoning	Not support	Support

Conclusion

- **Three semantic search models including basic model, extended and complete one.**
- **Combine text IR with semantic inference in the model.**
- **RBAC security ontology - OntoRBAC.**
- **A semantic search system (Onto-SSSE) is implemented based on the basic model.**

Future Work

We plan to achieve improvement in the following aspects:

- Perform search in a larger dataset.
- Implement the extended model and complete model.
- Improve the reasoning efficiency.

Thank you!