Automated Discovery, Categorization and Retrieval of Personalized Semantically Enriched E-learning Resources

Leyla Zhuhadar
Advisor: Prof. Olfa Nasraoui

1Knowledge Discovery and Web Mining Lab
Department of Computer Engineering and Computer Science
University of Louisville, USA.

Dissertation Defense: Doctor of Philosophy in Computer Science and Engineering
December 4, 2009
Outline

1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook
Outline

1. **Introduction**
   - “Big Issues” in Information Retrieval (IR)

2. **Research Question, Objective and Methodologies**
   - Our “Big Challenges”

3. **Overall Research Models**
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4. **Summary and Outlook**
The “Big Issues” in Information Retrieval.

- Performance: Efficient search and indexing (Bruce Croft, 2009) [2];
- Incorporating new data: Coverage and Freshness (Bruce Croft, 2009) [2];
- Scalability: Growing with data and users (Bruce Croft, 2009) [2];
- Adaptability: Tuning for applications and users (Bruce Croft, 2009)[2];
- Current problems: Information overload, keywords matching, ambiguity, handling evolution domain and users (Nasraoui: PKDD-2006-Invited-Talk) [6].
The “Big Issues” in Information Retrieval.

- Performance: Efficient search and indexing (Bruce Croft, 2009) [2];
- Incorporating new data: Coverage and Freshness (Bruce Croft, 2009) [2];
- Scalability: Growing with data and users (Bruce Croft, 2009) [2];
- Adaptability: Tuning for applications and users (Bruce Croft, 2009) [2];
- Current problems: Information overload, keywords matching, ambiguity, handling evolution domain and users (Nasraoui: PKDD-2006-Invited-Talk) [6].
The “Big Issues” in Information Retrieval.

- **Performance:** Efficient search and indexing (Bruce Croft, 2009) [2];
- **Incorporating new data:** Coverage and Freshness (Bruce Croft, 2009) [2];
- **Scalability:** Growing with data and users (Bruce Croft, 2009) [2];
- **Adaptability:** Tuning for applications and users (Bruce Croft, 2009)[2];
- **Current problems:** Information overload, keywords matching, ambiguity, handling evolution domain and users (Nasraoui: PKDD-2006-Invited-Talk) [6].
The “Big Issues” in Information Retrieval.

- **Performance**: Efficient search and indexing (Bruce Croft, 2009) [2];
- **Incorporating new data**: Coverage and Freshness (Bruce Croft, 2009) [2];
- **Scalability**: Growing with data and users (Bruce Croft, 2009) [2];
- **Adaptability**: Tuning for applications and users (Bruce Croft, 2009)[2];
- **Current problems**: Information overload, keywords matching, ambiguity, handling evolution domain and users (Nasraoui: PKDD-2006-Invited-Talk) [6].
The “Big Issues” in Information Retrieval.

- Performance: Efficient search and indexing (Bruce Croft, 2009) [2];
- Incorporating new data: Coverage and Freshness (Bruce Croft, 2009) [2];
- Scalability: Growing with data and users (Bruce Croft, 2009) [2];
- Adaptability: Tuning for applications and users (Bruce Croft, 2009)[2];
- Current problems: Information overload, keywords matching, ambiguity, handling evolution domain and users (Nasraoui: PKDD-2006-Invited-Talk) [6].
The main research question guiding this dissertation is whether it is feasible and beneficial to use different methodologies to design an information retrieval system for the E-learning domain, while still being able to retrieve personalized learning resources that are satisfactory and effective for the learner.
The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

- Metadata;
- Semantic knowledge representation;
- Natural language processing techniques;
- Clustering algorithms and;
- Visual representation.
Objective and Methodologies

The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

- Metadata;
- Semantic knowledge representation;
- Natural language processing techniques;
- Clustering algorithms and;
- Visual representation.
The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

- **Metadata;**
- **Semantic knowledge representation;**
- **Natural language processing techniques;**
- **Clustering algorithms and;**
- **Visual representation.**
The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

**Used Methodologies**

- Metadata;
- Semantic knowledge representation;
- Natural language processing techniques;
- Clustering algorithms and;
- Visual representation.
The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

- **Metadata**;
- **Semantic knowledge representation**;
- **Natural language processing techniques**;
- **Clustering algorithms and**;
- **Visual representation**.
The main methodology used in this research is to find empirical knowledge (by testing & evaluating) about the usefulness of using different methodologies to design an information retrieval system, either by using:

- Metadata;
- Semantic knowledge representation;
- Natural language processing techniques;
- Clustering algorithms and;
- Visual representation.
Proposed Architectural Design Methods.

- Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

**Architectural Design Methods**

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation model based on ontology.

\(^a\) [http://ocw.mit.edu/ocwweb/web/home/home/index.htm](http://ocw.mit.edu/ocwweb/web/home/home/index.htm)
Proposed Architectural Design Methods.

Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation ontology-based search model.

\(^a\) [http://ocw.mit.edu/OcwWeb/web/home/home/index.htm](http://ocw.mit.edu/OcwWeb/web/home/home/index.htm)
Proposed Architectural Design Methods.

- Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare;
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation ontology-based search model.

\[^a\] http://ocw.mit.edu/OcwWeb/web/home/home/index.htm
Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

**Architectural Design Methods**

- Metadata search model;
- **Personalized Semantic search model using ontologies and clustering techniques**;
- **Dual representation of the semantic user profile for personalized web search in an evolving domain**;
- An Augmented HyperManyMedia model with [MIT OpenCourseWare](http://ocw.mit.edu/OcwWeb/web/home/home/index.htm);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation model based on ontology.
Proposed Architectural Design Methods.

- Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation model based on ontology.

\(^a\) [http://ocw.mit.edu/OcwWeb/web/home/home/index.htm](http://ocw.mit.edu/OcwWeb/web/home/home/index.htm)
Proposed Architectural Design Methods.

- Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation model based on ontology.

\(^a\) [http://ocw.mit.edu/ocwweb/web/home/home/index.htm](http://ocw.mit.edu/ocwweb/web/home/home/index.htm)
Proposed Architectural Design Methods.

- Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- *Metadata search model;*
- *Personalized Semantic search model using ontologies and clustering techniques;*
- *Dual representation of the semantic user profile for personalized web search in an evolving domain;*
- *An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);*
- *Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering)*;
- *Cross-language ontology-based search model;*
- *Visual representation model based on ontology.*

\(^a\) [http://ocw.mit.edu/ocwweb/web/home/home/index.htm](http://ocw.mit.edu/ocwweb/web/home/home/index.htm)
Proposed Architectural Design Methods.

Using several Architectural Design Methods to build/implement IR Models and evaluate their: Scalability, Efficiency, and Usability.

Architectural Design Methods

- Metadata search model;
- Personalized Semantic search model using ontologies and clustering techniques;
- Dual representation of the semantic user profile for personalized web search in an evolving domain;
- An Augmented HyperManyMedia model with MIT OpenCourseWare\(^a\);
- Hybrid Recommender System (User Relevance Feedback, Collaborative Filtering);
- Cross-language ontology-based search model;
- Visual representation model based on ontology.

\(^a\) [http://ocw.mit.edu/ocweb/web/home/home/index.htm](http://ocw.mit.edu/ocweb/web/home/home/index.htm)
1. Introduction
   - “Big Issues” in Information Retrieval (IR)

2. Research Question, Objective and Methodologies
   - Our “Big Challenges”

3. Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4. Summary and Outlook
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
  - Design a Personalized Semantic search model;
  - Build a Cross-Language search model;
  - Utilize User Relevance Feedback to provide a personalized filtering method to users;
  - Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
  - Build a Cross-Language search model;
  - Utilize User Relevance Feedback to provide a personalized filtering method to users;
  - Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
How can We?

- Add Metadata to our resources to make them reachable;
- Extract the ontology of a specific domain to make it understandable from the semantic perspective;
- Re-model our system to be ready for the Semantic Web;
- Design a Personalized Semantic search model;
- Build a Cross-Language search model;
- Utilize User Relevance Feedback to provide a personalized filtering method to users;
- Use Collaborative Filtering techniques to take advantage of the similarity between users;
- And finally, map the domain ontology into a Visual search model.
Figure: Overall Research Models

- **Overall Modeling and Searching Techniques**
  - **Generic Search**
    - Using standard keyword Search (TF/IDF)
  - **Metadata Search**
    - Using Metadata Search Fields (Language, College, Course, etc.)
  - **Personalized Semantic Search**
    - Using OWL/RDF
  - **User Relevance Feedback**
    - Updating User Profiles based on his/her Relevance Feedback
  - **Recommender System**
    - Updating User Profiles based on similar Users
  - **Cross-Language Search**
    - Providing the user with MLIR (English/Spanish)
  - **Visual Search**
    - Using Prefuse to Visualize HyperManyMedia ontology

- **Vector Space Model (VSM)**
- **Metadata Techniques**
- **Ontology (Protegé)**
- **Rocchio’s Algorithm**
- **K-Nearest Neighbor & Fast XOR bit Operation Method**
- **Thesaurus-based Approach & Corpus-based Approach**
- **Prefuse Graph Model**
Outline

1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook
Phase I: Architectural Design/Implementation of Metadata Search Model.

Data about data: Metadata is valuable in the storage and retrieval of information (WordNet).

Definition

“There is a tremendous need for information retrieval methods that will exploit the core learning content and for metadata that can exploit further the learning value, (Sicilia, 2005)[8].”
Figure: Architectural Design/Implementation of Metadata Search Model
Phase I: Architectural Design/Implementation of Metadata Search Model.

Data about data: Metadata is valuable in the storage and retrieval of information (WordNet)

\[
\text{Score}(q, d) = \text{coord}(q, d) \times \text{queryNorm}(q) \times \sum(\text{tf}(t, \text{ind}) \times \text{idf}(t)^2 \times t \cdot \text{getBoost}() \times \text{norm}(t, d))
\]

- Features: \( \text{getBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) \)
- Weights: \( \text{getBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) \)

Modified Nutch Scoring Algorithm for Metadata

- Features: \( \text{ModifiedBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) + \varepsilon(\text{meta-data}) \)
- Weights: \( \text{ModifiedBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) + 5.0(\text{meta-data}) \)
Phase I: Architectural Design/Implementation of Metadata Search Model.

Data about data: Metadata is valuable in the storage and retrieval of information (WordNet)

\[
\text{Score}(q, d) = \text{coord}(q, d) \times \text{queryNorm}(q) \times \sum(\text{tf}(t, \text{index}) \times \text{idf}(t)^2 \times t \cdot \text{getBoost}() \times \text{norm}(t, d))
\]

- **Features:** \( \text{getBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) \)
- **Weights:** \( \text{getBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) \)

**Modified Nutch Scoring Algorithm for Metadata**

- **Features:** \( \text{ModifiedBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) + \epsilon(\text{meta-data}) \)
- **Weights:** \( \text{ModifiedBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) + 5.0(\text{meta-data}) \)
Phase I: Architectural Design/Implementation of Metadata Search Model.

Data about data: Metadata is valuable in the storage and retrieval of information (WordNet)

- **Score**\((q, d) = coord(q, d) \times queryNorm(q) \times \sum(tf(tind) \times idf(t)^2 \times t.getBoost() \times norm(t, d))$$
- Features: \(getBoost = \alpha(url) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title})$$
- Weights: \(getBoost = 4.0(url) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title})$$

Modified Nutch Scoring Algorithm for Metadata

- Features: \(\text{ModifiedBoost} = \alpha(url) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) + \varepsilon(\text{meta-data})$$
- Weights: \(\text{ModifiedBoost} = 4.0(url) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) + 5.0(\text{meta-data})$$
Phase I: Architectural Design/Implementation of Metadata Search Model.

Data about data: Metadata is valuable in the storage and retrieval of information (WordNet)

- Score: $\text{Score}(q, d) = \text{coord}(q, d) \times \text{queryNorm}(q) \times \sum(\text{tf}(t, \text{ind}) \times \text{idf}(t)^2 \times \text{t.getBoost()} \times \text{norm}(t, d))$
- Features: $\text{getBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title})$
- Weights: $\text{getBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title})$

Modified Nutch Scoring Algorithm for Metadata

- Features: $\text{ModifiedBoost} = \alpha(\text{url}) + \beta(\text{anchor}) + \gamma(\text{content}) + \delta(\text{title}) + \epsilon(\text{meta-data})$
- Weights: $\text{ModifiedBoost} = 4.0(\text{url}) + 2.0(\text{anchor}) + 1.0(\text{content}) + 1.0(\text{title}) + 5.0(\text{meta-data})$
Phase I: Evaluation of Metadata Search Model

- **Research Questions**
  1. Will there be an increase in precision when using the metadata search engine compared to the generic search engine?
  2. Will relevant documents be ranked higher when using the metadata search engine?

- **Selection of Queries**

- **Evaluation Measures**
  - \( \text{Precision} = \frac{\text{number of relevant documents}}{\text{number of retrieved documents}} \times 100 \)
  - \( \text{SEREET} = \frac{\sum Wi}{n} \times 100\% \)
  - \( Wi = \begin{cases} 
    100 & \text{if top of the rank} \\
    100 & \text{if no miss is found} \\
    p.p. - \frac{\text{#misses}}{\text{rank length}} \times 100\% & \text{or otherwise}
  \end{cases} \)
Phase I: Evaluation of Metadata Search Model

- **Research Questions**
  1. Will there be an increase in precision when using the metadata search engine compared to the generic search engine?
  2. Will relevant documents be ranked higher when using the metadata search engine?

- **Selection of Queries**

- **Evaluation Measures**
  - Precision: \( \text{Precision} = \frac{\text{number of relevant documents}}{\text{number of retrieved documents}} \times 100 \)
  - SEREET: \( \text{SEREET} = \frac{\sum W_i}{n} \times 100\% \)

  \[ W_i = \begin{cases} 
  100 & \text{if top of the rank} \\
  100 & \text{if no miss is found} \\
  p.p. - \frac{\# \text{misses}}{\text{rank length}} \times 100\% & \text{or otherwise}
  \end{cases} \]
Phase I: Evaluation of Metadata Search Model

Research Questions

1. Will there be an increase in precision when using the metadata search engine compared to the generic search engine?
2. Will relevant documents be ranked higher when using the metadata search engine?

Selection of Queries

Evaluation Measures

- **Precision** = \( \frac{\text{number of relevant documents}}{\text{number of retrieved documents}} \times 100 \)
- **SEREET** = \( \sum_{i=1}^{n} W_i \times 100\% \)

\[ W_i = \begin{cases} 
100 & \text{if top of the rank} \\
100 & \text{if no miss is found} \\
p \cdot p \cdot \frac{\text{misses}}{\text{rank length}} \times 100\% & \text{or otherwise}
\end{cases} \]
Phase I: Evaluation of Metadata Search Model

Research Questions

1. Will there be an increase in precision when using the metadata search engine compared to the generic search engine?
2. Will relevant documents be ranked higher when using the metadata search engine?

Selection of Queries

Evaluation Measures

\[ \text{Precision} = \frac{\text{number of relevant documents}}{\text{number of retrieved documents}} \times 100 \]

\[ \text{SEREEt} = \frac{\sum W_i}{n} \times 100\% \]

\[ W_i = \begin{cases} 
100 & \text{if top of the rank} \\
100 & \text{if no miss is found} \\
p \cdot p \cdot \frac{\# \text{misses}}{\text{rank length}} \times 100\% & \text{or otherwise}
\end{cases} \]
Table: Overall PRECISION for Metadata Search Model

<table>
<thead>
<tr>
<th>Query</th>
<th>Single-term</th>
<th>Two-term</th>
<th>Three-term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>0.619</td>
<td>0.717</td>
<td>0.851</td>
</tr>
<tr>
<td>Metadata</td>
<td>0.81</td>
<td>0.856</td>
<td>0.925</td>
</tr>
</tbody>
</table>

Figure: Overall PRECISION for Metadata Search Model
Table: Overall SEREET for Metadata Search Model

<table>
<thead>
<tr>
<th>Query</th>
<th>Single-term</th>
<th>Two-terms</th>
<th>Three-terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>0.597</td>
<td>0.684</td>
<td>0.834</td>
</tr>
<tr>
<td>Metadata</td>
<td>0.803</td>
<td>0.846</td>
<td>0.914</td>
</tr>
</tbody>
</table>

Figure: Overall SEREET for Metadata Search Model
Phase I: Evaluation Conclusion

We conclude that:

Evaluation of Metadata Search Model

- The Metadata search engine increases the precision across all numbers of terms used. This in turn answers our first research question: “Will there be an increase in precision when using the metadata search engine?”

- The Metadata search engine increases the ranking performance across all the numbers of terms that we used in queries. This, in turn, answers our second research question: “Will relevant documents be ranked higher when using a metadata search engine?”
**Phase I: Evaluation Conclusion**

- **We conclude that:**

**Evaluation of Metadata Search Model**

- The Metadata search engine increases the precision across all numbers of terms used. This in turn answers our first research question: “Will there be an increase in precision when using the metadata search engine?”

- The Metadata search engine increases the ranking performance across all the numbers of terms that we used in queries. This, in turn, answers our second research question: “Will relevant documents be ranked higher when using a metadata search engine?”
Phase I: Evaluation Conclusion

We conclude that:

Evaluation of Metadata Search Model

- The Metadata search engine increases the precision across all numbers of terms used. This in turn answers our first research question: “Will there be an increase in precision when using the metadata search engine?”

- The Metadata search engine increases the ranking performance across all the numbers of terms that we used in queries. This, in turn, answers our second research question: “Will relevant documents be ranked higher when using a metadata search engine?”
Outline

1. Introduction
   - “Big Issues” in Information Retrieval (IR)

2. Research Question, Objective and Methodologies
   - Our “Big Challenges”

3. Overall Research Models
   - Phase I: Metadata Search Model
   - **Phase II: Personalized Cluster-based Semantic Search Model**
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4. Summary and Outlook
Phase II: Our Assumptions are based on


- Clustering can improve Search Recall (Felman, 2006) [3].

- Clustering can improve Search Precision and may help in finding a user’s specific interest (Felman, 2006) [3].
Phase II: Our Assumptions are based on


- Clustering can improve Search Recall (Felman, 2006) [3].

- Clustering can improve Search Precision and may help in finding a user’s specific interest (Felman, 2006) [3].
Phase II: Our Assumptions are based on


- Clustering can improve Search Recall (Felman, 2006) [3].

- Clustering can improve Search Precision and may help in finding a user’s specific interest (Felman, 2006) [3].

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile—Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
- Best Cluster Mapping Algorithm for a Learner \(U_i\)—Algorithm 3.2: it measures the similarity \(Sim(D, CL_i)\)
- Re-ranking the Learner’s Search Results—Algorithm 3.3: Maps the ranked documents to semantic user profile
  - Each document \(d_i\), belonging to a semantic user profile, is assigned a priority ranking (\(\alpha = 5.0\))
  - Each document \(d_j\), belonging to the recommended cluster, is assigned a priority ranking (\(\beta = 3.0\))

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile–Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
  - Best Cluster Mapping Algorithm for a Learner $U_i$—Algorithm 3.2: it measures the similarity $Sim(D, CL_i)$
  - Re-ranking the Learner’s Search Results–Algorithm 3.3: Maps the ranked documents to semantic user profile
    - Each document $d_j$ belonging to a semantic user profile, is assigned a priority ranking ($\alpha = 5.0$)
    - Each document $d_j$ belonging to the recommended cluster, is assigned a priority ranking ($\beta = 3.0$)

Methodology/Implementation of Personalized Cluster-based Semantic Search Model:

- $R = \bigcup_{i=1}^{n} C_i$
- $C_i = \bigcup_{j=1}^{m} SC_{ji}$ if $C_i$ has subconcepts
- $C_i = \bigcup_{k=1}^{l} d_{ki}$ leaves

where $n$ = Number of concepts in the domain.

Each concept $C_i$ consists either of sub-concepts ($C_i = \bigcup_{j=1}^{m} SC_{ji}$) or of leaves ($C_i = \bigcup_{k=1}^{l} d_{ki}$)

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile–Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
- Best Cluster Mapping Algorithm for a Learner \( U_i \)–Algorithm 3.2: it measures the similarity \( \text{Sim}(D, CL_i) \)
- Re-ranking the Learner’s Search Results–Algorithm 3.3: Maps the ranked documents to semantic user profile
  - Each document \( d_j \), belonging to a semantic user profile, is assigned a priority ranking \( (\alpha = 5.0) \)
  - Each document \( d_j \) belonging to the recommended cluster, is assigned a priority ranking \( (\beta = 3.0) \)

\[ R = \bigcup_{i=1}^{n} C_i \]
\[
C_i = \bigcup_{j=1}^{m} SC_{ji} \quad \text{if } C_i \text{ has subconcepts}
\]
\[
C_i = \bigcup_{k=1}^{l} d_{ki} \quad \text{leaves}
\]

where \( n \) = Number of concepts in the domain.

Each concept \( C_i \) consists either of sub-concepts \( (C_i = \bigcup_{j=1}^{m} SC_{ji}) \) or of leaves \( (C_i = \bigcup_{k=1}^{l} d_{ki}) \)

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile–Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
- Best Cluster Mapping Algorithm for a Learner \( U_i \)–Algorithm 3.2: it measures the similarity \( \text{Sim}(D, CL_i) \)
- Re-ranking the Learner’s Search Results–Algorithm 3.3: Maps the ranked documents to semantic user profile
  - Each document \( d_j \) belonging to a semantic user profile, is assigned a priority ranking \( (\alpha = 5.0) \)
  - Each document \( d_j \) belonging to the recommended cluster, is assigned a priority ranking \( (\beta = 3.0) \)

R = \bigcup_{i=1}^{n} C_i

\begin{align*}
C_i &= \bigcup_{j=1}^{m} SC_{ji} \quad \text{if } C_i \text{ has subconcepts} \\
C_i &= \bigcup_{k=1}^{l} d_{ki} \quad \text{leaves}
\end{align*}

where \( n \) = Number of concepts in the domain.

Each concept \( C_i \) consists either of sub-concepts (\( C_i = \bigcup_{j=1}^{m} SC_{ji} \)) or of leaves (\( C_i = \bigcup_{k=1}^{l} d_{ki} \)).

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile–Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
- Best Cluster Mapping Algorithm for a Learner \( U_i \)–Algorithm 3.2: it measures the similarity \( \text{Sim}(D,CL_i) \)
- Re-ranking the Learner’s Search Results–Algorithm 3.3: Maps the ranked documents to semantic user profile
  - Each document \( d_{ij} \), belonging to a semantic user profile, is assigned a priority ranking (\( \alpha = 5.0 \))
  - Each document \( d_{ij} \) belonging to the recommended cluster, is assigned a priority ranking (\( \beta = 3.0 \))

\[
R = \bigcup_{i=1}^{n} C_i
\]

\[
C_i = \bigcup_{j=1}^{m} SC_{ji}
\]

if \( C_i \) has subconcepts

\[
\bigcup_{k=1}^{l} d_{ki}
\]

leaves

where \( n \) = Number of concepts in the domain.

Each concept \( C_i \) consists either of sub-concepts \( (C_i = \bigcup_{j=1}^{m} SC_{ji}) \) or of leaves \( (C_i = \bigcup_{k=1}^{l} d_{ki}) \)

Methodology/Implementation of Personalized Cluster-based Semantic Search Model

- Semantic Domain Structure (ontology of E-learning domain)
- Building Semantic User Profile–Algorithm 3.1: Bottom-up Pruning Algorithm: Building Learner’s Semantic Profile
- Best Cluster Mapping Algorithm for a Learner \( U_i \)–Algorithm 3.2: it measures the similarity \( \text{Sim}(D, CL_i) \)
- Re-ranking the Learner’s Search Results–Algorithm 3.3: Maps the ranked documents to semantic user profile
  - Each document \( d_i \) belonging to a semantic user profile, is assigned a priority ranking \( (\alpha = 5.0) \)
  - Each document \( d_i \) belonging to the recommended cluster, is assigned a priority ranking \( (\beta = 3.0) \)
Figure: Personalized Cluster-based Semantic Search Model
Figure: Semantic Domain Structure: Part of the tree structure generated from the OWL file
**Figure: Semantic User Profile**
Phase II: Evaluation of Personalized Cluster-based Semantic Search Model.
Main goal: to evaluate the effectiveness of re-ranking an information retrieval system using the Personalized Semantic representation of the User Profile, and Cluster Analysis.

**Research Question**

- Will personalized semantic search improve Recall and Precision?

**Corpus**

- 10 concepts (colleges), 28 subconcepts (courses)
- Dataset (a total of 2,812 documents)
Phase II: Evaluation of Personalized Cluster-based Semantic Search Model.

Main goal: to evaluate the effectiveness of re-ranking an information retrieval system using the Personalized Semantic representation of the User Profile, and Cluster Analysis.

- **Research Question**
  - Will personalized semantic search improve Recall and Precision?

- **Corpus**
  - 10 concepts (colleges), 28 subconcepts (courses)
  - Dataset (a total of 2,812 documents)
Phase II: Evaluation of Personalized Cluster-based Semantic Search Model.

Main goal: to evaluate the effectiveness of re-ranking an information retrieval system using the Personalized Semantic representation of the User Profile, and Cluster Analysis.

- **Research Question**
  - Will personalized semantic search improve Recall and Precision?

- **Corpus**
  - 10 concepts (colleges), 28 subconcepts (courses)
  - Dataset (a total of 2,812 documents)
Phase II: Evaluation Measures.

- Evaluation Measures
  - Clustering:
    - Total corpus consisting of around 2,812 documents (lectures)
    - Experiment with partitional algorithms, direct K-way clustering (similar to K-means), and repeated bisection or Bisecting K-Means with all criterion functions.
    - Repeated each clustering algorithm with all possible combinations of clustering criterion functions for different numbers of clusters: 20, 25, 30, 35, 40, 45, 50.
    - \(E(S_r) = -\frac{1}{\log q} \sum_{i=1}^{q} \frac{n_i}{n_r} \log \frac{n_i}{n_r}\) (Zhao, Y. and Karypis, G., Criterion functions for document clustering: Experiments and analysis, 2001)[9]
    - \(E(T) = \Sigma_{r=1}^{p} \frac{1}{p} E(S_r)\) (Zhao, Y. and Karypis, G., Criterion functions for document clustering: Experiments and analysis, 2001)[9]
  - Top-n-Recall and Top-n-Precision:
    - \(Top - n \text{ Recall} = \frac{\text{number of relevant retrieved documents within top n results}}{\text{total number of relevant documents}}\)
    - \(Top - n \text{ Precision} = \frac{\text{number of relevant retrieved documents within top n results}}{n}\)
    - Average Percentage of Improvement in Top-n Recall
    - Average Percentage of Improvement in Top-n Precision
    - 10 profiles with three sizes of queries (1, 2, and 3 keywords)
Phase II: Evaluation Measures.

**Evaluation Measures**

- **Clustering:**
  - Total corpus consisting of around 2,812 documents (lectures)
  - Experiment with partitional algorithms, direct K-way clustering (similar to K-means), and repeated bisection or Bisecting K-Means with all criterion functions.
  - Repeated each clustering algorithm with all possible combinations of clustering criterion functions for different numbers of clusters: 20, 25, 30, 35, 40, 45, 50.
  - \[ E(S_r) = -\frac{1}{\log q} \sum_{i=1}^{q} \frac{n^i}{n^r} \log \frac{n^i}{n^r} \] (Zhao, Y. and Karypis, G., Criterion functions for document clustering: Experiments and analysis, 2001)[9]
  - \[ E(T) = \sum_{r=1}^{p} \frac{1}{p} E(S_r) \] (Zhao, Y. and Karypis, G., Criterion functions for document clustering: Experiments and analysis, 2001)[9]

- **Top-n-Recall and Top-n-Precision:**
  - \[ \text{Top} - n \text{ Recall} = \frac{\text{number of relevant retrieved documents within top n results}}{\text{total number of relevant documents}} \]
  - \[ \text{Top} - n \text{ Precision} = \frac{\text{number of relevant retrieved documents within top n results}}{n} \]
  - Average Percentage of Improvement in Top-n Recall
  - Average Percentage of Improvement in Top-n Precision
  - 10 profiles with three sizes of queries (1, 2, and 3 keywords)
Phase II: Evaluation Results.
2,812 lectures, 7 different formats, a total of ~20,000

Evaluation Results

- Graph-partitioning produced the best clustering results
- K=35 clusters and the lowest entropy.
- Confusion matrix (41 misclassified documents out of 2812)

Personalized vs. Non-Personalized: Top-n-Recall and Top-n-Precision Results

- Improvement in precision that varies between 5-25%.
- Improvement is noticeable between the top-30 and top-50 for single-keyword and two-keywords queries.
- Improvement in Recall between top-20 and top-40.
Phase II: Evaluation Results.
2,812 lectures, 7 different formats, a total of ~20,000

- Evaluation Results
  - Graph-partitioning produced the best clustering results
  - K=35 clusters and the lowest entropy.
  - Confusion matrix (41 misclassified documents out of 2812)

- Personalized vs. Non-Personalized: Top-n-Recall and Top-n-Precision Results
  - Improvement in precision that varies between 5-25%.
  - Improvement is noticeable between the top-30 and top-50 for single-keyword and two-keywords queries.
  - Improvement in Recall between top-20 and top-40.
Figure: Average Percentage of Improvement in Top-n Precision and Top-n Recall
Phase II: Evaluation Conclusion

2,812 lectures, 7 different formats, a total of ~20,000

We conclude that:

Evaluation of Personalized Cluster-based Semantic Search Model

Changing the ranking mechanism of a search engine according to the semantic user profile improves the Average Percentage of Top-n Recall and the Average Percentage of Top-n Precision, which answers the research question in phase II.
Phase II: Evaluation Conclusion

2,812 lectures, 7 different formats, a total of ~20,000

We conclude that:

Evaluation of Personalized Cluster-based Semantic Search Model

Changing the ranking mechanism of a search engine according to the semantic user profile improves the Average Percentage of Top-n Recall and the Average Percentage of Top-n Precision, which answers the research question in phase II.
1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - **Phase III: Dual Representation of the Semantic User Profile**
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook

Dissertation Defense: Doctor of Philosophy in CSE–Leyla Zhuhadar
Automated DCR of Personalized Semantic IR System
Phase III: Dual Representation of the Semantic User Profile for Personalized Web Search in an Evolving Domain

“Our assumption that the users’ interests must converge after a period of time.”

- **Research Questions**
  1. Will the users’ interests converge after a period of time?
  2. If the users’ interests converge, can we estimate the time frame of the convergence?

- **Evaluation Measures**
  - Convergence: Find the time frame where the users’ interests converge.
Phase III: Dual Representation of the Semantic User Profile for Personalized Web Search in an Evolving Domain

“Our assumption that the users’ interests must converge after a period of time.”

Research Questions

1. Will the users’ interests converge after a period of time?
2. If the users’ interests converge, can we estimate the time frame of the convergence?

Evaluation Measures

Convergence: Find the time frame where the users’ interests converge.
Phase III: Dual Representation of the Semantic User Profile for Personalized Web Search in an Evolving Domain

“Our assumption that the users’ interests must converge after a period of time.”

Research Questions

1. Will the users’ interests converge after a period of time?
2. If the users’ interests converge, can we estimate the time frame of the convergence?

Evaluation Measures

Convergence: Find the time frame where the users’ interests converge.
Phase III: Architectural Design/Implementation.

2,812 lectures, 7 different formats, a total of ~20,000

Update Phase II Architecture:

1. Building the Lower-level of Semantic User Profile
2. Building the Higher-level of Semantic User Profile

Dual Representation of the Semantic User Profile

- Algorithm 3.4: Tracking the User’s History of Interests
- Algorithm 3.5: Detecting Shifts in the User’s Interests
- The shift of interests affects two parts of the system:
  1. The cluster to profile ontology mapping
  2. Changing the semantic user profiles
Phase III: Architectural Design/Implementation.
2,812 lectures, 7 different formats, a total of ~20,000

Update Phase II Architecture:
1. Building the Lower-level of Semantic User Profile
2. Building the Higher-level of Semantic User Profile

Dual Representation of the Semantic User Profile

- Algorithm 3.4: Tracking the User’s History of Interests
- Algorithm 3.5: Detecting Shifts in the User’s Interests
- The shift of interests affects two parts of the system:
  1. The cluster to profile ontology mapping
  2. Changing the semantic user profiles
Phase III: Architectural Design/Implementation.
2,812 lectures, 7 different formats, a total of ~20,000

Update Phase II Architecture:
1. Building the Lower-level of Semantic User Profile
2. Building the Higher-level of Semantic User Profile

Dual Representation of the Semantic User Profile

- Algorithm 3.4: Tracking the User’s History of Interests
- Algorithm 3.5: Detecting Shifts in the User’s Interests

The shift of interests affects two parts of the system:
1. The cluster to profile ontology mapping
2. Changing the semantic user profiles
Phase III: Architectural Design/Implementation.

2,812 lectures, 7 different formats, a total of ~20,000

**Update Phase II Architecture:**
1. Building the Lower-level of Semantic User Profile
2. Building the Higher-level of Semantic User Profile

**Dual Representation of the Semantic User Profile**

- Algorithm 3.4: Tracking the User’s History of Interests
- Algorithm 3.5: Detecting Shifts in the User’s Interests
- The shift of interests affects two parts of the system:
  1. The cluster to profile ontology mapping
  2. Changing the semantic user profiles
Figure: Dual Representation of the Semantic User Profiles for Personalized Web Search in an Evolving Domain
Figure: User Interests vs. Sub-concept for 10 User Profiles
**Figure:** Convergence of User Profiles
We conclude that:

- The users’ interests converge after a period of time; this answers our first research question in Phase III.
- The time frame of convergence = 1 month, which answers our second research question in Phase III.
Phase III: Evaluation Conclusion

We conclude that:

**Dual Representation of the Semantic User Profiles in an Evolving Domain**

- The users’ interests converge after a period of time; this answers our first research question in Phase III.
- The time frame of convergence = 1 month, which answers our second research question in Phase III.
Phase III: Evaluation Conclusion

- We conclude that:

**Dual Representation of the Semantic User Profiles in an Evolving Domain**

- The users’ interests converge after a period of time; this answers our first research question in Phase III.
- The time frame of convergence = 1 month, which answers our second research question in Phase III.
Introduction

“Big Issues” in Information Retrieval (IR)

Research Question, Objective and Methodologies

Our “Big Challenges”

Overall Research Models

Phase I: Metadata Search Model
Phase II: Personalized Cluster-based Semantic Search Model
Phase III: Dual Representation of the Semantic User Profile
Phase IV: Augmenting HyperManyMedia Repository with MIT
Phase V: Recommender Search Engine (User-centered Approach)
Phase VI: Multi-Language Ontology-based Search Engine
Phase VII: Visual Ontology-based Search Model

Summary and Outlook
Phase IV: Augmenting HyperManyMedia Repository with MIT Resources

Table: Summary of HyperManyMedia Resources

<table>
<thead>
<tr>
<th>Total # of colleges= 11</th>
<th>Total # of courses = 64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # of WKU courses= 27</td>
<td>Total # of MIT courses = 37</td>
</tr>
<tr>
<td>Total # of English courses= 45</td>
<td>Total # of Spanish courses = 19</td>
</tr>
<tr>
<td></td>
<td>Total # of Lectures = 7,424</td>
</tr>
<tr>
<td></td>
<td>Total # of LO = 51,968</td>
</tr>
</tbody>
</table>
Figure: HyperManyMedia User Interface (UI)
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
  - Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
  - Phase VI: Multi-Language Ontology-based Search Engine
  - Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Add the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- **Adding the MIT OpenCourseWare**
- **Extending the Ontology**
- **Changing:** Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
  - Building an Extended Ontology Structure
  - Clustering/Adding the descriptive features
  - Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/ Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Methodology for Augmenting HyperManyMedia Search Engines

- Adding the MIT OpenCourseWare
- Extending the Ontology
- Changing: Generic search, Metadata search, and Semantic search
- Phase V: Recommender Semantic Search Engine (User Relevance Feedback and Collaborative Filtering)
- Phase VI: Multi-Language Ontology-based Search Engine
- Phase VII: Visual Ontology-based Search Engine

Methodology for Augmenting HyperManyMedia Search Engines

- Collecting similar courses/lectures located in MIT
- Parsing the Metadata/Creating local plugins
- Building an Extended Ontology Structure
- Clustering/Adding the descriptive features
- Crawling and indexing the HyperManyMedia platform
- Collecting users profiles
- Re-evaluating the System
Phase IV: Augmenting Semantic Search

- Constructing the Extended HyperManyMedia Ontology: Using Protégé to define the following Entities: (has_College, has_Course, has_Language, has_Lecture, has_Professor, and sub_Class_Of)

**Example**

**The characteristics of Entity = has_College are:**
- **Description:** College
- **Equivalentclasses:** Colegio
- **Superclasses:** Thing
- **Members:** Accounting, Architecture_and_Manufacturing, Biology, etc.

**Example**

**Equivalent classes:** Equivalent classes equal to ≡ relation, to mention some of these entities (College ≡ Colegio, Engineering ≡ Ingenieria, English ≡ Ingles, ..., Social Work ≡ Trabajo Social, Chemistry ≡ Quimica, etc.)

**Example**

Sub_Class_Of: Related to the hierarchy design of our domain: <Cluster_descriptive_features is – a sub_Class_Of Lecture>, <Lecture is – a sub_Class_Of Course>, <Course is – a sub_Class_Of College>, etc.
Phase IV: Augmenting Semantic Search

- **Constructing the Extended HyperManyMedia Ontology:** Using Protégé to define the following Entities: *(has_College, has_Course, has_Language, has_Lecture, has_Professor, and sub_Class_Of)*

**Example**

The characteristics of *Entity = has_College* are: **Description:** College, **Equivalentclasses:** Colegio, **Superclasses:** Thing, **Members:** Accounting, Architecture_and_Manufacturing, Biology, etc.

**Example**

**Equivalent classes:** *Equivalent classes* equal to ≡ relation, to mention some of these entities (College ≡ Colegio, Engineering ≡ Ingenieria, English ≡ Ingles, ..., Social Work≡Trabajo Social, Chemistry≡Quimica, etc.)

**Example**

Sub_Class_Of: Related to the hierarchy design of our domain: <Cluster_descriptive features is – a sub_Class_Of Lecture>, <Lecture is – a sub_Class_Of Course>, <Course is – a sub_Class_Of College>, etc.
Phase IV: Augmenting Semantic Search

Constructing the Extended HyperManyMedia Ontology: Using Protégé to define the following Entities: (has_College, has_Course, has_Language, has_Lecture, has_Professor, and sub_Class_Of)

Example

The characteristics of Entity = has_College are: Description: College, Equivalent classes: Colegio, Superclasses: Thing, Members: Accounting, Architecture_and_Manufacturing, Biology, etc.

Example

Equivalent classes: Equivalent classes equal to ≡ relation, to mention some of these entities (College ≡ Colegio, Engineering ≡ Ingenieria, English ≡ Ingles,..., Social Work ≡ Trabajo Social, Chemistry ≡ Quimica, etc.)

Example

Sub_Class_Of: Related to the hierarchy design of our domain: <Cluster_descriptive_features is – a sub_Class_Of Lecture>, <Lecture is – a sub_Class_Of Course>, <Course is – a sub_Class_Of College>, etc.
Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto\(^1\)
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

\(^1\)http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview
Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto\(^1\)
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

\(^1\)http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview
Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto\(^1\)
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

\(^1\)http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview
Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto\(^1\)
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

\(^1\) [http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview](http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview)
Phase IV: Clustering/Text Analysis for the Semantic Search

- **Clustering English Documents:**
  - Using package Cluto\(^1\)
  - English corpus consisting of 4,888 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

- **Clustering Spanish Documents:**
  - Using package Cluto
  - English corpus consisting of 2,536 documents
  - Repeat each clustering algorithm with all possible combinations of clustering criterion functions for different number of clusters: 20, 25, 30, 35, 40, 45, 50

\(^1\) [http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview](http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview)
Figure: Semantic Search
Phase IV: Evaluation of Augmented HyperManyMedia Search Engine

- **Research Questions**
  1. Will there be an improvement in Top-n-Recall when using the Metadata search engine compared to the Generic search engine?
  2. Will there be an improvement in Top-n-Precision when using the Metadata search engine compared to the Generic search engine?
  3. Will there be an improvement in Top-n-Recall when using the Semantic search engine compared to the Metadata and Generic search engine?
  4. Will there be an improvement in Top-n-Precision when using the Semantic search engine compared to the Metadata and Generic search engine?
Phase IV: Evaluation of Augmented HyperManyMedia Search Engine

- **Research Questions**
  1. Will there be an improvement in Top-n-Recall when using the Metadata search engine compared to the Generic search engine?
  2. Will there be an improvement in Top-n-Precision when using the Metadata search engine compared to the Generic search engine?
  3. Will there be an improvement in Top-n-Recall when using the Semantic search engine compared to the Metadata and Generic search engine?
  4. Will there be an improvement in Top-n-Precision when using the Semantic search engine compared to the Metadata and Generic search engine?
Phase IV: Evaluation of Augmented HyperManyMedia Search Engine

Research Questions

1. Will there be an improvement in Top-n-Recall when using the Metadata search engine compared to the Generic search engine?
2. Will there be an improvement in Top-n-Precision when using the Metadata search engine compared to the Generic search engine?
3. Will there be an improvement in Top-n-Recall when using the Semantic search engine compared to the Metadata and Generic search engine?
4. Will there be an improvement in Top-n-Precision when using the Semantic search engine compared to the Metadata and Generic search engine?
Phase IV: Evaluation of Augmented HyperManyMedia Search Engine

**Research Questions**

1. Will there be an improvement in Top-n-Recall when using the Metadata search engine compared to the Generic search engine?
2. Will there be an improvement in Top-n-Precision when using the Metadata search engine compared to the Generic search engine?
3. Will there be an improvement in Top-n-Recall when using the Semantic search engine compared to the Metadata and Generic search engine?
4. Will there be an improvement in Top-n-Precision when using the Semantic search engine compared to the Metadata and Generic search engine?
Phase IV: Clustering Results.

Evaluation Results

- **Clustering English Documents (4,888)**: The best clustering method we found for the English corpus which produced the highest \( \text{Purity} = 0.959 \) and with the lowest \( \text{Entropy} = 0.05 \) was the Agglomerative Method, with **Number of Clusters = 38**, using **Clustering Criterion Function = II**, as and **Cosine Similarity Measures** as **inter-object similarity measure**.

- **Clustering Spanish Documents (2,536)**: The best clustering method we found for the Spanish corpus, which produced the highest \( \text{Purity} = 0.927 \) and with the lowest \( \text{Entropy} = 0.140 \), was the Agglomerative Method, with **Number of Clusters = 50**, using **Clustering Criterion Function = II** and **Cosine Similarity Measures** as **inter-object similarity measure**.
Phase IV: Clustering Results.

**Evaluation Results**

- **Clustering English Documents (4,888)**: The best clustering method we found for the English corpus which produced the highest $Purity = 0.959$ and with the lowest $Entropy = 0.05$ was the Agglomerative Method, with $Number of Clusters = 38$, using $Clustering Criterion Function = II$, as and $Cosine Similarity Measures$ as *inter-object similarity measure*.

- **Clustering Spanish Documents (2,536)**: The best clustering method we found for the Spanish corpus, which produced the highest $Purity = 0.927$ and with the lowest $Entropy = 0.140$, was the Agglomerative Method, with $Number of Clusters = 50$, using $Clustering Criterion Function = II$ and $Cosine Similarity Measures$ as *inter-object similarity measure*. 
Figure: Top-n-Recall (Generic vs. Metadata vs. Semantic) for 1,825 user profiles
Figure: Top-n-Precision (Generic vs. Metadata vs. Semantic) for 1,825 user profiles
We conclude that:

- The Metadata search engine shows an improvement in Recall compared to the Generic search engine in all Top-n intervals, which answers our first research question in Phase IV.
- The Metadata search engine shows an improvement in Precision compared to the Generic search engine in all Top-n intervals, which answers our second research question in Phase IV.
- Our assumption that the Recall results using the Semantic search will outperform the Metadata results failed. We noticed that the Recall in the Metadata search engine outperformed both Generic and Semantic results, which answers our third research question in Phase IV.
- Our assumption that Precision results using the Semantic search will outperform the Metadata results was marginally supported. We noticed that the Precision results in the Semantic search engine outperform (from Top-10 to Top-60) the Generic and Metadata search engines, then the Precision dropped in the Semantic search as compared to the Metadata search, which answers our fourth research question in Phase IV.
Phase IV: Evaluation Conclusion
2,812 lectures, 7 different formats, a total of ~20,000

We conclude that:

- The Metadata search engine shows an improvement in Recall compared to the Generic search engine in all Top-n intervals, which answers our first research question in Phase IV.
- The Metadata search engine shows an improvement in Precision compared to the Generic search engine in all Top-n intervals, which answers our second research question in Phase IV.
- Our assumption that the Recall results using the Semantic search will outperform the Metadata results failed. We noticed that the Recall in the Metadata search engine outperformed both Generic and Semantic results, which answers our third research question in Phase IV.
- Our assumption that Precision results using the Semantic search will outperform the Metadata results was marginally supported. We noticed that the Precision results in the Semantic search engine outperform (from Top-10 to Top-60) the Generic and Metadata search engines, then the Precision dropped in the Semantic search as compared to the Metadata search, this answers our fourth research question in Phase IV.
We conclude that:

1. The Metadata search engine shows an improvement in Recall compared to the Generic search engine in all Top-n intervals, which answers our first research question in Phase IV.

2. The Metadata search engine shows an improvement in Precision compared to the Generic search engine in all Top-n intervals, which answers our second research question in Phase IV.

3. Our assumption that the Recall results using the Semantic search will outperform the Metadata results failed. We noticed that the Recall in the Metadata search engine outperformed both Generic and Semantic results, which answers our third research question in Phase IV.

4. Our assumption that Precision results using the Semantic search will outperform the Metadata results was marginally supported. We noticed that the Precision results in the Semantic search engine outperform (from Top-10 to Top-60) the Generic and Metadata search engines, then the Precision dropped in the Semantic search as compared to the Metadata search, this answers our fourth research question in Phase IV.
Phase IV: Evaluation Conclusion

2,812 lectures, 7 different formats, a total of ~20,000

We conclude that:

Evaluation of Augmented HyperManyMedia Search Engine

1. The Metadata search engine shows an improvement in Recall compared to the Generic search engine in all Top-n intervals, which answers our first research question in Phase IV.

2. The Metadata search engine shows an improvement in Precision compared to the Generic search engine in all Top-n intervals, which answers our second research question in Phase IV.

3. Our assumption that the Recall results using the Semantic search will outperform the Metadata results failed. We noticed that the Recall in the Metadata search engine outperformed both Generic and Semantic results, which answers our third research question in Phase IV.

4. Our assumption that Precision results using the Semantic search will outperform the Metadata results was marginally supported. We noticed that the Precision results in the Semantic search engine outperform (from Top-10 to Top-60) the Generic and Metadata search engines, then the Precision dropped in the Semantic search as compared to the Metadata search, this answers our fourth research question in Phase IV.
Phase IV: Evaluation Conclusion

2,812 lectures, 7 different formats, a total of ~20,000

We conclude that:

Evaluation of Augmented HyperManyMedia Search Engine

1. The Metadata search engine shows an improvement in Recall compared to the Generic search engine in all Top-n intervals, which answers our first research question in Phase IV.

2. The Metadata search engine shows an improvement in Precision compared to the Generic search engine in all Top-n intervals, which answers our second research question in Phase IV.

3. Our assumption that the Recall results using the Semantic search will outperform the Metadata results failed. We noticed that the Recall in the Metadata search engine outperformed both Generic and Semantic results, which answers our third research question in Phase IV.

4. Our assumption that Precision results using the Semantic search will outperform the Metadata results was marginally supported. We noticed that the Precision results in the Semantic search engine outperform (from Top-10 to Top-60) the Generic and Metadata search engines, then the Precision dropped in the Semantic search as compared to the Metadata search, this answers our fourth research question in Phase IV.
Outline

1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - **Phase V: Recommender Search Engine (User-centered Approach)**
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook
Phase V: Recommender Search Engine

Preliminary Finding: The architecture of phases II and III has the following weaknesses:

- Proficiency (time)
- Scalability (number of users)

Building User Profiles was based on the log file (this process is time consuming and can only be done offline).

Updating Users Profiles was controlled by the system administrator and not by the users themselves.

The previous design was limited to few users.

Tracking the user profile via log files was not considered as an accurate procedure, especially if we compare it to a database system that collects users profiles, user updates, and user interactions with the system.
Phase V: Recommender Search Engine

- Preliminary Finding: The architecture of phases II and III has the following weaknesses:
  - Proficiency (time)
  - Scalability (number of users)

Recommender Search Engine

- Building User Profiles was based on the log file (this process is time consuming and can only be done offline).
- Updating Users Profiles was controlled by the system administrator and not by the users themselves.
- The previous design was limited to few users.
- Tracking the user profile via log files was not considered as an accurate procedure, especially if we compare it to a database system that collects users profiles, user updates, and user interactions with the system.
Phase V: Recommender Search Engine

- Preliminary Finding: The architecture of phases II and III has the following weaknesses:
  - Proficiency (time)
  - Scalability (number of users)

Recommender Search Engine

1. Building User Profiles was based on the log file (this process is time consuming and can only be done offline).
2. Updating Users Profiles was controlled by the system administrator and not by the users themselves
3. The previous design was limited to few users.
4. Tracking the user profile via log files was not considered as an accurate procedure, especially if we compare it to a database system that collects users profiles, user updates, and user interactions with the system.
Phase V: Recommender Search Engine

Preliminary Finding: The architecture of phases II and III has the following weaknesses:

- Proficiency (time)
- Scalability (number of users)

Recommender Search Engine

- Building User Profiles was based on the log file (this process is time consuming and can only be done offline).
- Updating Users Profiles was controlled by the system administrator and not by the users themselves.
- The previous design was limited to few users.
- Tracking the user profile via log files was not considered as an accurate procedure, especially if we compare it to a database system that collects users profiles, user updates, and user interactions with the system.
Phase V: Recommender Search Engine

Preliminary Finding: The architecture of phases II and III has the following weaknesses:

- Proficiency (time)
- Scalability (number of users)

Recommender Search Engine

1. Building User Profiles was based on the log file (this process is time consuming and can only be done offline).
2. Updating Users Profiles was controlled by the system administrator and not by the users themselves.
3. The previous design was limited to few users.
4. Tracking the user profile via log files was not considered as an accurate procedure, especially if we compare it to a database system that collects users profiles, user updates, and user interactions with the system.

**Semantic User Profile: Algorithm 3.7** (Pruning Algorithm: Building Semantic User Profile)

- Assume $P_i$ is a profile for a user $i$ that consists of all the documents $\sum_{k=1}^l d_{ki}$ which belong to the user’s interests [college/course(s)/lecture(s)]
- $\sum_{j=1}^m d'_{ji}$ represents all the documents belonging to [college(s)/course(s)/lecture(s)] in which the user $i$ is not interested.
- We represent the user profile as $P_i = \alpha \sum_{k=1}^l d_{ki} + \beta \sum_{j=1}^m d'_{ji}$, where $d_i$ is a vector representing document $i$, and $\alpha, \beta$ are parameters that control the tradeoff between the two document types.
- We consider two methods of weighting:
  1. **Exploitation Method** ($\alpha = 1$ and $\beta = 0$): A complete filtering to all documents $\sum_{j=1}^m d'_{ji}$ which are not related to the user profile.
  2. **Exploration Method** ($\alpha = 0.8$ and $\beta = 0.2$)[2]: A partial filtering to all documents $\sum_{j=1}^m d'_{ji}$ which are not related to the user profile.
- Given a query $q$, for a user profile $P_i$, we modified the boosting in the semantic search by adding a user profile plugin: $P'_i = \gamma(P_i) = \gamma(\alpha \sum_{k=1}^l d_{ki} + \beta \sum_{j=1}^m d'_{ji})$
- where $P'_i$ represents the modified user profile, $\alpha, \beta$ are parameters that control the tradeoff between the two document types (as we discussed above), and $\gamma$ is the boosting factor.
Phase V: Updating User Profile Based on Interests (*User Relevance Feedback*).

**Update Semantic User Profile:**
We employ a variant of *Rocchio’s Algorithm* as follows:

\[
P_i'' = P_i' - \frac{1}{|\text{Non-Rel}|} \sum_{d_i \in \text{Non-Rel}} d_i
\]
Figure: Updating User Profile Based on Interests (User Relevance Feedback)
Phase V: Recommendations Based on Collaborative Filtering

- Two methods were applied on the HyperManyMedia:
  1. The K-Nearest Neighbors classifier
  2. The user-to-user fast XOR bit operation method
Phase V: Recommendations Based on Collaborative Filtering

- Two methods were applied on the HyperManyMedia:
  1. The K-Nearest Neighbors classifier
  2. The user-to-user fast XOR bit operation method
Phase V: Recommendations Based on Collaborative Filtering

- Two methods were applied on the HyperManyMedia:
  1. The K-Nearest Neighbors classifier
  2. The user-to-user fast XOR bit operation method
Phase V: Implementing Collaborative Filtering Method II: Fast XOR bit Operation Method

Figure: Fast XOR bit Operations Method

- 11 Colleges (64 courses), $2^6 = 64$ distinct codes, 8 inter-connected cube would represents the whole domain (64 course)
- $0100 \rightarrow 1001$ has distance 3 (red path)
- $0110 \rightarrow 1110$ has distance 1 (blue path)
- Algorithm 3.9: Represents the pruning methodology
- Algorithm 3.10: Represents a recommendation based on Fast XOR bit Operations Method given to a user $U_i$ requesting Collaborative Filtering.
Phase V: Evaluation of Recommender System Model

- **Research Questions**
  1. Will there be an improvement in Top-n-Recall when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?
  2. Will there be an improvement in Top-n-Precision when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?
  3. Will there be an improvement in Top-n-Recall when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?
  4. Will there be an improvement in Top-n-Precision when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?
  5. Will there be an improvement in Top-n-Recall and Top-n-Precision when using Personalized Semantic search using Collaborative Filtering compared to the Personalized Semantic search engine?
Phase V: Evaluation of Recommender System Model

Research Questions

1. Will there be an improvement in Top-n-Recall when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

2. Will there be an improvement in Top-n-Precision when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

3. Will there be an improvement in Top-n-Recall when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

4. Will there be an improvement in Top-n-Precision when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

5. Will there be an improvement in Top-n-Recall and Top-n-Precision when using Personalized Semantic search using Collaborative Filtering compared to the Personalized Semantic search engine?
Phase V: Evaluation of Recommender System Model

Research Questions

1. Will there be an improvement in Top-n-Recall when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

2. Will there be an improvement in Top-n-Precision when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

3. Will there be an improvement in Top-n-Recall when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

4. Will there be an improvement in Top-n-Precision when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

5. Will there be an improvement in Top-n-Recall and Top-n-Precision when using Personalized Semantic search using Collaborative Filtering compared to the Personalized Semantic search engine?
Phase V: Evaluation of Recommender System Model

Research Questions

1. Will there be an improvement in Top-n-Recall when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

2. Will there be an improvement in Top-n-Precision when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

3. Will there be an improvement in Top-n-Recall when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

4. Will there be an improvement in Top-n-Precision when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

5. Will there be an improvement in Top-n-Recall and Top-n-Precision when using Personalized Semantic search using Collaborative Filtering compared to the Personalized Semantic search engine?
Phase V: Evaluation of Recommender System Model

Research Questions

1. Will there be an improvement in Top-n-Recall when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

2. Will there be an improvement in Top-n-Precision when using the Personalized Semantic search engine compared to Non-Personalized Semantic search engine?

3. Will there be an improvement in Top-n-Recall when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

4. Will there be an improvement in Top-n-Precision when using Personalized Semantic search with Relevance Feedback compared to the Personalized Semantic search engine?

5. Will there be an improvement in Top-n-Recall and Top-n-Precision when using Personalized Semantic search using Collaborative Filtering compared to the Personalized Semantic search engine?
Figure: Top-n-Recall (Semantic vs. Personalized Semantic vs. Personalized Semantic with Relevance Feedback)
**Figure:** Top-n-Precision (Semantic vs. Personalized Semantic vs. Personalized Semantic with Relevance Feedback)
Figure: Top-n-Recall (Personalized Semantic Search with Collaborative Filtering vs. Non-Collaborative Filtering) [dataset of only 127 user profiles to evaluate the Top-n-Recall. These profiles belong to the Mathematics College. In the original dataset (1,825 user profiles), they present a subset ranging from profile #411 to profile #537]
Figure: Top-n-Precision (Personalized Semantic with Collaborative Filtering vs. Non-Collaborative Filtering) [dataset of only 127 user profiles to evaluate the Top-n-Recall. These profiles belong to the Mathematics College. In the original dataset (1,825 user profiles), they present a subset ranging from profile #411 to profile #537]
Phase V: Evaluation Conclusion

- **We conclude that:**

  Evaluation of Recommender System Model

  1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.

  2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.

  3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.

  4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answers our fourth research question in Phase V.

  5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answers our fifth research question in Phase V.
Phase V: Evaluation Conclusion

We conclude that:

1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.

2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.

3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.

4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answers our fourth research question in Phase V.

5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answers our fifth research question in Phase V.
We conclude that:

1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.
2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.
3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.
4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answer our fourth research question in Phase V.
5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answer our fifth research question in Phase V.
We conclude that:

### Evaluation of Recommender System Model

1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.

2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.

3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.

4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answers our fourth research question in Phase V.

5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answers our fifth research question in Phase V.
Phase V: Evaluation Conclusion

We conclude that:

Evaluation of Recommender System Model

1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.

2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.

3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.

4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answers our fourth research question in Phase V.

5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answers our fifth research question in Phase V.
Phase V: Evaluation Conclusion

We conclude that:

Evaluation of Recommender System Model

1. Personalized Semantic search shows an improvement in Recall compared to the Semantic search engine in all Top-n intervals, which answers our first research question in Phase V.

2. Personalized Semantic search shows an improvement in Precision compared to the Semantic search engine in all Top-n intervals, which answers our second research question in Phase V.

3. Personalized Semantic search with Relevance Feedback shows a slight improvement in Recall compared to the Personalized Semantic search with Relevance Feedback, which answers our third research question in Phase V.

4. Personalized Semantic search with Relevance Feedback shows a high improvement in Precision compared to the Personalized Semantic search with Relevance Feedback, which answers our fourth research question in Phase V.

5. Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration, which answers our fifth research question in Phase V.
Outline

1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - **Phase VI: Multi-Language Ontology-based Search Engine**
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook
Multi-Language Ontology-based Search Engine

- **MLIR approach falls into a Domain Specific Retrieval (E-learning)**
  - Synergistic approach between
    1. Thesaurus-based Approach and;
    2. Query translation approach (Douglas, 1996)[7]
Multi-Language Ontology-based Search Engine

- **MLIR approach falls into a Domain Specific Retrieval (E-learning)**
- Synergistic approach between
  1. Thesaurus-based Approach and;
  2. Query translation approach (Douglas, 1996)[7]
Multi-Language Ontology-based Search Engine

- MLIR approach falls into a Domain Specific Retrieval (E-learning)
- Synergistic approach between
  - Thesaurus-based Approach and;
  - Query translation approach (Douglas, 1996)[7]
Multi-Language Ontology-based Search Engine

- **MLIR approach falls into a Domain Specific Retrieval (E-learning)**
- Synergistic approach between
  1. Thesaurus-based Approach and;
  2. Query translation approach (Douglas, 1996)[7]
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Thesaurus-based Approach

- Simple bilingual ontology thesaurus listing of terms, phrases, concepts, and subconcepts;
- Terminology used to capture the HyperManyMedia domain and presented in two languages (English and Spanish).
- Query translation approach: whenever a user submits a query in the semantic search interface, the following two parallel processes occur:
  - All relevant documents to the query term will be retrieved, and the ranking of those documents will be based on the scoring algorithm in Phase I
  - An automatic semantic mapping between the query term and the HyperManyMedia ontology, which is resident in memory, if the query term is a part of the HyperManyMedia ontology; the information retrieval system will automatically present two semantic entities:
    - All the subconcepts related to this query term in both languages (English and Spanish)
    - Synonym to the query term in the alternative language
Corpus-based Approach: Term Vector Translation

A user submits a query in one of the languages, English or Spanish, and clicks on cross-language translation. If the query contains part of our indexed translated terms, the search engine does the following:

1. Translate the query $q$ to the alternative language $q'$, as shown in Algorithm 3.11.
2. Use the Vector Space Model to calculate the dot product between the translated query and the documents in the HyperManyMedia repository, after substituting each $q$ to $q'$ to retrieve relevant documents and ranks them based on the score.
Corpus-based Approach: Term Vector Translation

A user submits a query in one of the languages, English or Spanish, and clicks on cross-language translation. If the query contains part of our indexed translated terms, the search engine does the following:

1. Translate the query $q$ to the alternative language $q'$, as shown in Algorithm 3.11.
2. Use the Vector Space Model to calculate the dot product between the translated query and the documents in the HyperManyMedia repository, after substituting each $q$ to $q'$ to retrieve relevant documents and ranks them based on the score.
A user submits a query in one of the languages, English or Spanish, and clicks on cross-language translation. If the query contains part of our indexed translated terms, the search engine does the following:

1. Translate the query $q$ to the alternative language $q'$, as shown in **Algorithm 3.11**.
2. Use the Vector Space Model to calculate the dot product between the translated query and the documents in the HyperManyMedia repository, after substituting each $q$ to $q'$ to retrieve relevant documents and ranks them based on the score.
Phase VI: Evaluation of Cross-Language Search Model

**Research Question**

Will there be a difference in Top-n-Recall and Top-n-Precision when we Cross from the Spanish Language to the English language vs. from the English Language to the Spanish?

**Evaluation Measures**

- We use Top-n-Recall and Top-n-Precision
Phase VI: Evaluation of Cross-Language Search Model

Research Question

Will there be a difference in Top-n-Recall and Top-n-Precision when we Cross from the Spanish Language to the English language vs. from the English Language to the Spanish?

Evaluation Measures

- We use Top-n-Recall and Top-n-Precision
Figure: Top-n-Recall on multi-levels for Cross-Language Search Engine
Figure: Top-n-Precision on multi-levels for Cross-Language Search Engine
We conclude that:

Cross-Language Search Engine

The Cross-language search engine performs better when we cross from the Spanish language to the English language in the top-n-recall and Top-n-Precision, which answers our second research question in Phase VI.
Phase VI: Evaluation Conclusion

We conclude that:

Cross-Language Search Engine

- The Cross-language search engine performs better when we cross from the Spanish language to the English language in the top-n-recall and Top-n-Precision, which answers our second research question in Phase VI.
Phase VI: Evaluation Conclusion

- We conclude that:

Cross-Language Search Engine
- The Cross-language search engine performs better when we cross from the Spanish language to the English language in the top-n-recall and Top-n-Precision, which answers our second research question in Phase VI.
Phase VI: Evaluation Conclusion

Fact

The following reasons have influenced the results:

- English courses have been indexed and boosted in multiple stages during the design of the platform (during the last two years).
- Adding the Spanish courses was done during a very short period of time; thus we have not been able to add sophisticated tagging to these resources, because of the time constraints.
- The ontology relationships between the two languages need to be logically improved using a higher level of interrelationship between entities and concepts.
Fact

The following reasons have influenced the results:

- **English courses have been indexed and boosted in multiple stages during the design of the platform (during the last two years).**

- **Adding the Spanish courses was done during a very short period of time; thus we have not been able to add sophisticated tagging to these resources, because of the time constraints.**

- **The ontology relationships between the two languages need to be logically improved using a higher level of interrelationship between entities and concepts.**
Phase VI: Evaluation Conclusion

Fact

The following reasons have influenced the results:

- **English courses have been indexed and boosted in multiple stages during the design of the platform (during the last two years).**

- **Adding the Spanish courses was done during a very short period of time; thus we have not been able to add sophisticated tagging to these resources, because of the time constraints.**

- **The ontology relationships between the two languages need to be logically improved using a higher level of interrelationship between entities and concepts.**
Phase VI: Evaluation Conclusion

Fact

**The following reasons have influenced the results:**

- *English courses have been indexed and boosted in multiple stages during the design of the platform (during the last two years).*

- *Adding the Spanish courses was done during a very short period of time; thus we have not been able to add sophisticated tagging to these resources, because of the time constraints.*

- *The ontology relationships between the two languages need to be logically improved using a higher level of interrelationship between entities and concepts.*
Outline

1 Introduction
   - “Big Issues” in Information Retrieval (IR)

2 Research Question, Objective and Methodologies
   - Our “Big Challenges”

3 Overall Research Models
   - Phase I: Metadata Search Model
   - Phase II: Personalized Cluster-based Semantic Search Model
   - Phase III: Dual Representation of the Semantic User Profile
   - Phase IV: Augmenting HyperManyMedia Repository with MIT
   - Phase V: Recommender Search Engine (User-centered Approach)
   - Phase VI: Multi-Language Ontology-based Search Engine
   - Phase VII: Visual Ontology-based Search Model

4 Summary and Outlook
How is it possible to visualize an ontology graph which represents knowledge and reasoning of a massive, ambiguous, and vast document set using minimum vocabulary?

- Our Assumptions are based on

**Used Methodologies**

- **Zipf’s laws**—the *Principle of Least Effort* (Manning, 1999)[5]: building a small set of vocabulary that represents the whole domain of our repository.

- **Collocation Concept** (Manning, 1999): our ontology not only consists of nouns, but also of compound phrases (e.g., Introduction to Literature: as a course name).
How is it possible to visualize an ontology graph which represents knowledge and reasoning of a massive, ambiguous, and vast document set using minimum vocabulary?

Our Assumptions are based on

Used Methodologies

- **Zipf’s laws**— the *Principle of Least Effort* (Manning, 1999)[5]: building a small set of vocabulary that represents the whole domain of our repository.

- **Collocation Concept** (Manning, 1999): our ontology not only consists of nouns, but also of compound phrases (e.g., Introduction to Literature: as a course name).
How is it possible to visualize an ontology graph which represents knowledge and reasoning of a massive, ambiguous, and vast document set using minimum vocabulary?

Our Assumptions are based on

Used Methodologies

- **Zipf’s laws**– the *Principle of Least Effort* (Manning, 1999)[5]: building a small set of vocabulary that represents the whole domain of our repository.

- **Collocation Concept** (Manning, 1999): our ontology not only consists of nouns, but also of compound phrases (e.g., Introduction to Literature: as a course name).
Phase VII: Implementation of Visual-base Search Model

Definition

A formal context, is defined [4] SG (Simple Conceptual Graph) as a triple \((D, \delta, \alpha)\) where \(D\) is a set of objects and \(\delta\) is a set of attributes and \(\alpha\) defines the relationship between \(D\) and \(\delta\). For example, let us build a model \((D, \delta, \alpha)\) satisfying G, which is in our case represents a semantic representation of the HyperManyMedia domain.
Figure: Illustrating the scenario of representing a simple conceptual graph
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

  The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  - Left Mouse Button Click on a Sector
  - Right Mouse Button Click on a Sector
  - Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  - Left Mouse Button Click on a Sector
  - Right Mouse Button Click on a Sector
  - Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  - Left Mouse Button Click on a Sector
  - Right Mouse Button Click on a Sector
  - Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  - Left Mouse Button Click on a Sector
  - Right Mouse Button Click on a Sector
  - Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- Usability Test consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- Functionality Test
  1. Left Mouse Button Click on a Sector
  2. Right Mouse Button Click on a Sector
  3. Double-Click on a Sector
Usability Test consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

The test covered three levels of testing:
1. based on the hierarchical level of the ontology domain;
2. based on the English resources in each level;
3. based on the Spanish resources in each level.

Functionality Test
- Left Mouse Button Click on a Sector
- Right Mouse Button Click on a Sector
- Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  1. Left Mouse Button Click on a Sector
  2. Right Mouse Button Click on a Sector
  3. Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

Usability Test consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

The test covered three levels of testing:

1. based on the hierarchical level of the ontology domain;
2. based on the English resources in each level;
3. based on the Spanish resources in each level.

Functionality Test

- Left Mouse Button Click on a Sector
- Right Mouse Button Click on a Sector
- Double-Click on a Sector
Phase VII: Evaluation of Visual Ontology-based Search Model

- **Usability Test** consists of evaluating each concept and subconcept presented in the visual interface. (refer to Table 4.50)

- The test covered three levels of testing:
  1. based on the hierarchical level of the ontology domain;
  2. based on the English resources in each level;
  3. based on the Spanish resources in each level.

- **Functionality Test**
  1. Left Mouse Button Click on a Sector
  2. Right Mouse Button Click on a Sector
  3. Double-Click on a Sector
Phase VII: Left mouse Button Click on a Sector (level=1)

**Figure:** One Level Filtering of the query “Engineering”

![Filtering Diagram](image-url)
Figure: Two Levels Filtering of the query “Engineering”
Phase VII: Left mouse Button Click on a Sector (level=3)

**Figure:** Three Levels Filtering of the query “Engineering”
Phase VII: Double-click on a Sector

Figure: Double-click on the “Engineering” Sector
Phase VII: Right mouse Button Click on a Sector

Figure: Right clicking on the “Engineering” Sector
Phase VII: Active Visual Search

- A user types the first three letters “Eng”:
  - shows 186 concepts retrieved by only typing “E”;
  - shows 32 concepts retrieved by typing “En”;
  - shows 4 concepts retrieved by typing “Eng”

1. Monarchy in England (Lecture:SubSubSubConcept level)
2. English (College:SubConcept level)
3. Engineering (College:SubConcept level)
4. Introduction to Computers and Engineering (Course:SubSubConcept level)
Phase VII: Active Visual Search

- A user types the three letters "Eng":
  - shows 186 concepts retrieved by only typing "E";
  - shows 32 concepts retrieved by typing "En";
  - shows 4 concepts retrieved by typing "Eng"

1. Monarchy in England (Lecture:SubSubSubConcept level)
2. English (College:SubConcept level)
3. Engineering (College:SubConcept level)
4. Introduction to Computers and Engineering (Course:SubSubConcept level)
Phase VII: Active Visual Search

A user types the three letters “Eng”:

- shows 186 concepts retrieved by only typing “E”;
- shows 32 concepts retrieved by typing “En”;
- shows 4 concepts retrieved by typing “Eng”

1. Monarchy in England (Lecture:SubSubSubConcept level)
2. English (College:SubConcept level)
3. Engineering (College:SubConcept level)
4. Introduction to Computers and Engineering (Course:SubSubConcept level)
Phase VII: Active Visual Search

A user type the three letters “Eng”:

- shows 186 concepts retrieved by only typing “E”;
- shows 32 concepts retrieved by typing “En”;
- shows 4 concepts retrieved by typing “Eng”

1. Monarchy in England (Lecture:SubSubSubConcept level)
2. English (College:SubConcept level)
3. Engineering (College:SubConcept level)
4. Introduction to Computers and Engineering (Course:SubSubConcept level)
Phase VI: Active Visual Search “E”

Figure: Visual Search for “E”
**Phase VI: Active Visual Search “En”**

**Figure:** Visual Search for “En”
Phase VI: Active Visual Search “Eng”

Figure: Visual Search for “Eng”
Phase VI: Evaluation Results

- Precision is used to evaluate the accuracy of this phase. Building this visual interface is based on an ontology. Table 4.51 presents the visual concepts that have been tested.

- 20 concepts and subconcepts randomly have been selected to test the visual interface.
Phase VI: Evaluation Results

- Precision is used to evaluate the accuracy of this phase. Building this visual interface is based on an ontology. Table 4.51 presents the visual concepts that have been tested.

- 20 concepts and subconcepts randomly have been selected to test the visual interface.
Phase VI: Evaluation Results cont.

We conclude that:

- When the visual term is a concept, the precision is very low.
- Whereas, when the visual term is a subconcept, the precision is very high.
- On average the Top-20-precision with random sampling of concepts/subconcepts was 0.637.
We conclude that:

Visual Ontology-based Search Model

- When the visual term is a concept, the precision is very low.
- Whereas, when the visual term is a subconcept, the precision is very high.
- On average the Top-20-precision with random sampling of concepts/subconcepts was 0.637
We conclude that:

Visual Ontology-based Search Model

- When the visual term is a concept, the precision is very low.
- Whereas, when the visual term is a subconcept, the precision is very high.
- On average the **Top-20-precision** with random sampling of concepts/subconcepts was 0.637
Phase VI: Evaluation Results cont.

- **We conclude that:**

**Visual Ontology-based Search Model**

- When the visual term is a concept, the precision is very low.
- Whereas, when the visual term is a subconcept, the precision is very high.
- On average the **Top-20-precision** with random sampling of concepts/subconcepts was **0.637**
Comparison between All the Models

- The Metadata search engine outperformed the Generic search engine in each interval (Top-10, Top20, Top-30, ..., Top-100) of Top-n-Recall. However, the best recall obtained using the Metadata search engine was 0.750 (at Top-100) compared to 0.31 (at Top-100) in Generic search engine.

- The Metadata search engine outperformed the Generic search engine in each interval (Top-10, Top20, Top-30, ..., Top-100) of Top-n-Precision. However, the best precision obtained using the Metadata search engine was 0.69 (at top-100) compared to 0.16 (at top-40) in the Generic search engine. It is noticeable that after augmenting the system with external resources, the precision of the Generic search engine became very low for each Top-n interval.

- Metadata search outperforms Generic and Semantic search in almost each interval (Top-10, Top20, Top-30, ..., Top-100) of Top-n-Recall. However, the results of the Metadata and Semantic search are very close (Semantic recall only outperformed Metadata recall at Top-60 and at Top-80). However, The recall performance of Semantic Search dropped to 0.50 (at Top-100) compared to 0.69 (at Top-100) in Metadata search and to 0.31 (at Top-100) in Generic search.
The Metadata search engine outperformed the Generic search engine in each interval (Top-10, Top20, Top-30,..., Top-100) of Top-n-Recall. However, the best recall obtained using the Metadata search engine was 0.750 (at Top-100) compared to 0.31 (at Top-100) in Generic search engine.

The Metadata search engine outperformed the Generic search engine in each interval (Top-10, Top20, Top-30,..., Top-100) of Top-n-Precision. However, the best precision obtained using the Metadata search engine was 0.69 (at top-100) compared to 0.16 (at top-40) in the Generic search engine. It is noticeable that after augmenting the system with external resources, the precision of the Generic search engine became very low for each Top-n interval.

Metadata search outperforms Generic and Semantic search in almost each interval (Top-10, Top20, Top-30,..., Top-100) of Top-n-Recall. However, the results of the Metadata and Semantic search are very close (Semantic recall only outperformed Metadata recall at Top-60 and at Top-80). However, The recall performance of Semantic Search dropped to 0.50 (at Top-100) compared to 0.69 (at Top-100) in Metadata search and to 0.31 (at Top-100) in Generic search.
Comparison between All the Models

- **The Metadata search engine** outperformed the **Generic search engine** in each interval (Top-10, Top-20, Top-30, ..., Top-100) of Top-n-Recall. However, the best recall obtained using the **Metadata search engine** was 0.750 (at Top-100) compared to 0.31 (at Top-100) in **Generic search engine**.

- **The Metadata search engine** outperformed the **Generic search engine** in each interval (Top-10, Top-20, Top-30, ..., Top-100) of Top-n-Precision. However, the best precision obtained using the **Metadata search engine** was 0.69 (at top-100) compared to 0.16 (at top-40) in the **Generic search engine**. It is noticeable that after augmenting the system with external resources, the precision of the **Generic search engine** became very low for each Top-n interval.

- **Metadata search** outperforms **Generic** and **Semantic search** in almost each interval (Top-10, Top-20, Top-30, ..., Top-100) of Top-n-Recall. However, the results of the **Metadata** and **Semantic search** are very close (**Semantic recall only outperformed Metadata recall at Top-60 and at Top-80**). However, **The recall performance of Semantic Search dropped to 0.50 (at Top-100) compared to 0.69 (at Top-100) in Metadata search and to 0.31 (at Top-100) in Generic search.**
Comparison between All the Models Cont.

- **Semantic search** outperformed *Generic search* and *Metadata search in Top-n-Precision* from *Top-10* until *Top-60*. However, the precision of the *Metadata search* got better than the *Semantic search* at *Top-80* and *Top100*. The best precision for the *Semantic search* was at *Top-40* = 0.66 compared to 0.7 (at *Top-100*) in the *Metadata search* and to 0.31 (at *Top-100*) in the *Generic search*.

- The **Personalized Semantic search** outperformed the *Metadata search* and the *Non-Personalized Semantic search*. In addition, using the *User Relevance Feedback* brought slightly better recall results.

- The **Personalized Semantic search** outperformed the *Metadata search* and the *Non-Personalized Semantic Search* in all interval. In addition, using *User Relevance Feedback* brought very high precision results.

- The **Personalized Semantic search with Collaborative Filtering** shows an improvement in Recall and Precision compared to the *Personalized Semantic search without Collaboration*.

- **Recall and Precision** in the *Cross-Language search* performed better when we crossed from the *Spanish language* to the *English*.
Comparison between All the Models Cont.

- *Semantic search* outperformed *Generic search* and *Metadata search* in Top-n-Precision from Top-10 until Top-60. However, the precision of the *Metadata search* got better than the *Semantic search* at Top-80 and Top100. The best precision for the *Semantic search* was at Top-40 = 0.66 compared to 0.7 (at Top-100) in the *Metadata search* and to 0.31 (at Top-100) in the *Generic search*.

- *The Personalized Semantic search* outperformed the *Metadata search* and the *Non-Personalized Semantic search*. In addition, using the *User Relevance Feedback* brought slightly better recall results.

- *The Personalized Semantic search* outperformed the *Metadata Search* and the *Non-Personalized Semantic Search* in all interval. In addition, using *User Relevance Feedback* brought very high precision results.

- *The Personalized Semantic search with Collaborative Filtering* shows an improvement in Recall and Precision compared to the *Personalized Semantic search without Collaboration*.

- *Recall and Precision* in the *Cross-Language search* performed better when we crossed from the *Spanish language* to the *English*. 
Comparison between All the Models Cont.

- **Semantic search** outperformed **Generic search** and **Metadata search** in Top-n-Precision from Top-10 until Top-60. However, the precision of the **Metadata search** got better than the **Semantic search** at Top-80 and Top-100. The best precision for the **Semantic search** was at Top-40 = 0.66 compared to 0.7 (at Top-100) in the **Metadata search** and to 0.31 (at Top-100) in the **Generic search**.

- The **Personalized Semantic search** outperformed the **Metadata search** and the **Non-Personalized Semantic search**. In addition, using the **User Relevance Feedback** brought slightly better recall results.

- The **Personalized Semantic search** outperformed the **Metadata Search** and the **Non-Personalized Semantic Search** in all interval. In addition, using **User Relevance Feedback** brought very high precision results.

- The **Personalized Semantic search with Collaborative Filtering** shows an improvement in Recall and Precision compared to the **Personalized Semantic search without Collaboration**.

- **Recall and Precision** in the **Cross-Language search** performed better when we crossed from the **Spanish language** to the **English**.
Comparison between All the Models Cont.

• **Semantic search** outperformed **Generic search** and **Metadata search in Top-n-Precision** from **Top-10** until **Top-60**. However, the precision of the **Metadata search** got better than the **Semantic search** at **Top-80** and **Top100**. The best precision for the **Semantic search** was at **Top-40 = 0.66** compared to **0.7 (at Top-100)** in the **Metadata search** and to **0.31 (at Top-100)** in the **Generic search**.

• **The Personalized Semantic search** outperformed the **Metadata search** and the **Non-Personalized Semantic search**. In addition, using the **User Relevance Feedback** brought slightly better recall results.

• **The Personalized Semantic search** outperformed the **Metadata Search** and the **Non-Personalized Semantic Search** in all interval. In addition, using **User Relevance Feedback** brought very high precision results.

• **The Personalized Semantic search with Collaborative Filtering shows an improvement in Recall and Precision compared to the Personalized Semantic search without Collaboration.**

• **Recall and Precision in the Cross-Language search** performed better when we crossed from the **Spanish language to the English.**
Comparison between All the Models Cont.

- **Semantic search** outperformed **Generic search** and **Metadata search** in Top-n-Precision from Top-10 until Top-60. However, the precision of the **Metadata search** got better than the **Semantic search** at Top-80 and Top-100. The best precision for the **Semantic search** was at Top-40 = 0.66 compared to 0.7 (at Top-100) in the **Metadata search** and to 0.31 (at Top-100) in the **Generic search**.

- **The Personalized Semantic search** outperformed the **Metadata search** and the **Non-Personalized Semantic search**. In addition, using the **User Relevance Feedback** brought slightly better recall results.

- **The Personalized Semantic search** outperformed the **Metadata Search** and the **Non-Personalized Semantic Search** in all interval. In addition, using **User Relevance Feedback** brought very high precision results.

- **The Personalized Semantic search** with **Collaborative Filtering** shows an improvement in **Recall and Precision** compared to the **Personalized Semantic search** without Collaboration.

- **Recall and Precision** in the **Cross-Language search** performed better when we crossed from the **Spanish language** to the **English**.

---

Dissertation Defense: Doctor of Philosophy in CSE–Leyla Zhuhadar

Automated DCR of Personalized Semantic IR System
**Figure:** Usage of the Multi-Search Engines (logs of 1,825 User Profiles)

- **Visual Search**: 28%
- **Generic Search**: 3%
- **Metadata Search**: 10%
- **Semantic Search**: 13%
- **Personalized Semantic Search**: 20%
- **Personalized Search with User Relevance Feedback**: 7%
- **Cross-Language Search**: 11%
- **Collaborative Filtering**: 8%
Figure: HyperManyMedia Search Engine Performance
The main goals of the *HyperManyMedia* system as implementation

**The suggested architectures:**

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the HyperManyMedia system as implementation

The suggested architectures:

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the HyperManyMedia repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the *HyperManyMedia* system as implementation

The suggested architectures:

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the *HyperManyMedia* system as implementation

The suggested architectures:

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- **Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare**
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the *HyperManyMedia* system as implementation

**The suggested architectures:**

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the HyperManyMedia system as implementation

The suggested architectures:

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the HyperManyMedia repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the *HyperManyMedia* system as implementation

**The suggested architectures:**
- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
The main goals of the *HyperManyMedia* system as implementation

The suggested architectures:

- Provide the learner with a metadata search engine
- Deliver a personalized semantic information retrieval system to learners using ontologies.
- Provide a dual representation of the semantic user profile for personalized web search in an evolving domain
- Augment the *HyperManyMedia* repository with external open source resources from MIT OpenCourseWare
- Provide each learner with user relevance feedback and collaborative filtering
- Provide the learner with a multilingual Ontology-based search engine
- Provide the learner with a visual search engine
- Be an open source repository for online learning
Contribution to the State of the Art

- This approach is significantly different than the currently used in the development of information retrieval systems:
  - Utilizes ontologies as models to provide semantic information.
  - Uses two different type of ontologies, a global ontology model that represents the whole E-learning domain, and a user ontology model that represents the learner profile.
  - The implementation of the ontology models is separate from the design and implementation of the information retrieval system (contributes to the State of the Art by providing an architecture that enables a generic approach that is application independent and reusable, and is therefore less costly to develop).
Contribution to the State of the Art

- This approach is significantly different than the currently used in the development of information retrieval systems:
  - Utilizes ontologies as models to provide semantic information.
    - Uses two different type of ontologies, a global ontology model that represents the whole E-learning domain, and a user ontology model that represents the learner profile.
    - The implementation of the ontology models is separate from the design and implementation of the information retrieval system (contributes to the State of the Art by providing an architecture that enables a generic approach that is application independent and reusable, and is therefore less costly to develop).
This approach is significantly different than the currently used in the development of information retrieval systems:

- Utilizes ontologies as models to provide semantic information.
- Uses two different type of ontologies, a global ontology model that represents the whole E-learning domain, and a user ontology model that represents the learner profile.
- The implementation of the ontology models is separate from the design and implementation of the information retrieval system (contributes to the State of the Art by providing an architecture that enables a generic approach that is application independent and reusable, and is therefore less costly to develop).
Contribution to the State of the Art

- This approach is significantly different than the currently used in the development of information retrieval systems:

  - Utilizes ontologies as models to provide semantic information.
  - Uses two different type of ontologies, a global ontology model that represents the whole E-learning domain, and a user ontology model that represents the learner profile.
  - The implementation of the ontology models is separate from the design and implementation of the information retrieval system (contributes to the State of the Art by providing an architecture that enables a generic approach that is application independent and reusable, and is therefore less costly to develop).
Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- Reuse of the Domain Ontology for the Personalized Semantic Search Model.
- Reuse of the Domain Ontology for the User Relevance Feedback Model.
- Reuse of the Domain Ontology for the Collaborative Filtering Model.
- Reuse of the Domain Ontology for the Cross-Language Search Model.
- Reuse of the Domain Ontology for the Visual Ontology-based Search Model.
Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- **Reuse of the Domain Ontology for the Personalized Semantic Search Model.**
- Reuse of the Domain Ontology for the User Relevance Feedback Model.
- Reuse of the Domain Ontology for the Collaborative Filtering Model.
- Reuse of the Domain Ontology for the Cross-Language Search Model.
- Reuse of the Domain Ontology for the Visual Ontology-based Search Model.
Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- **Reuse of the Domain Ontology for the Personalized Semantic Search Model.**
- **Reuse of the Domain Ontology for the User Relevance Feedback Model.**
- **Reuse of the Domain Ontology for the Collaborative Filtering Model.**
- **Reuse of the Domain Ontology for the Cross-Language Search Model.**
- **Reuse of the Domain Ontology for the Visual Ontology-based Search Model.**
Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- **Reuse of the Domain Ontology for the Personalized Semantic Search Model.**
- **Reuse of the Domain Ontology for the User Relevance Feedback Model.**
- **Reuse of the Domain Ontology for the Collaborative Filtering Model.**
- **Reuse of the Domain Ontology for the Cross-Language Search Model.**
- **Reuse of the Domain Ontology for the Visual Ontology-based Search Model.**
Contribution to the State of the Art Cont.

Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- **Reuse of the Domain Ontology for the Personalized Semantic Search Model.**
- **Reuse of the Domain Ontology for the User Relevance Feedback Model.**
- **Reuse of the Domain Ontology for the Collaborative Filtering Model.**
- **Reuse of the Domain Ontology for the Cross-Language Search Model.**
- **Reuse of the Domain Ontology for the Visual Ontology-based Search Model.**
Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets:

- Reuse of the Domain Ontology for the Personalized Semantic Search Model.
- Reuse of the Domain Ontology for the User Relevance Feedback Model.
- Reuse of the Domain Ontology for the Collaborative Filtering Model.
- Reuse of the Domain Ontology for the Cross-Language Search Model.
- Reuse of the Domain Ontology for the Visual Ontology-based Search Model.
Finally, we added to the State of Art a novel approach for recommendation, we named this approach “User-to-User Fast XOR bit Operations Method”. This method is considered as a Collaborative Filtering approach that provides a fast recommendation based on community of users; the strength of this approach relies on its simplicity.
Finally, we added to the State of Art a novel approach for recommendation, we named this approach “User-to-User Fast XOR bit Operations Method”.

This method is considered as a Collaborative Filtering approach that provides a fast recommendation based on community of users; the strength of this approach relies on its simplicity.
The influence of our research on the State of the Art has been witnessed by its contribution to over twenty publications in international conferences, journals and book chapters.
The influence of our research on the State of the Art has been witnessed by its contribution to over twenty publications in international conferences, journals and book chapters.
Future Work/Post Doctoral

- Post Doctoral goals are:
  - Extend the multilingual information retrieval system by using an automatic construction of the thesaurus
  - Using computational linguistics to build the thesauri
  - In the domain of Semantic Web, “Linked Data” is the right place to extend this research. Linked Data is a project directed by Christian Bizer, Tom Heath and Tim Berners-Lee [1]. Based on the growth of Linked Data (In 2009, Linked Data extended into 6.7 billion RDF triples and around 149 million RDF links) might be the future of the Semantic Web.
Post Doctoral goals are:

**Future Work**

- Extend the multilingual information retrieval system by using an automatic construction of the thesaurus
  - Using computational linguistics to build the thesauri
  - In the domain of Semantic Web, “Linked Data” is the right place to extend this research. Linked Data is a project directed by Christian Bizer, Tom Heath and Tim Berners-Lee [1]. Based on the growth of Linked Data (In 2009, Linked Data extended into 6.7 billion RDF triples and around 149 million RDF links) might be the future of the Semantic Web.
Future Work/Post Doctoral

- Post Doctoral goals are:

Future Work

- **Extend the multilingual information retrieval system by using an automatic construction of the thesaurus**

- **Using computational linguistics to build the thesauri**

- In the domain of Semantic Web, “Linked Data” is the right place to extend this research. Linked Data is a project directed by Christian Bizer, Tom Heath and Tim Berners-Lee [1]. Based on the growth of Linked Data (In 2009, Linked Data extended into 6.7 billion RDF triples and around 149 million RDF links) might be the future of the Semantic Web.
Post Doctoral goals are:

Future Work

- Extend the multilingual information retrieval system by using an automatic construction of the thesaurus
- Using computational linguistics to build the thesauri
- In the domain of Semantic Web, “Linked Data” is the right place to extend this research. Linked Data is a project directed by Christian Bizer, Tom Heath and Tim Berners-Lee [1]. Based on the growth of Linked Data (In 2009, Linked Data extended into 6.7 billion RDF triples and around 149 million RDF links) might be the future of the Semantic Web.
Figure: LinkedOpenData Datasets on the Web: July 2009
Research Links

Important Links:

- HyperManyMedia: \(^a\)
- HyperManyMedia Ontology (~40,000) line of code \(^b\)
- HyperManyMedia Visual Ontology (~167,000) line of code \(^c\)
- Complete Evaluation Results: \(^d\)
- KNN in Nutch \(^e\)

\(^a\) http://hypermangmedia.wku.edu
\(^b\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl
\(^c\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml
\(^d\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/
\(^e\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/
**Important Links:**

- **HyperManyMedia:**
  - HyperManyMedia Ontology (~40,000) line of code
  - HyperManyMedia Visual Ontology (~167,000) line of code
  - Complete Evaluation Results:
  - KNN in Nutch

---

- **HyperManyMedia:** [http://hypermanymedia.wku.edu](http://hypermanymedia.wku.edu)
- **Evaluation Results/semanticnew.owl:** [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl)
- **Evaluation Results/VisualOntologyGraph.xml:** [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml)
- **Evaluation Results/WekaInNutch/:** [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/)
Introduction

Research Question, Objective and Methodologies

Overall Research Models

Summary and Outlook

---

**Important Links:**

- HyperManyMedia: \[^a\]
- HyperManyMedia Ontology (~40,000) line of code \[^b\]
- HyperManyMedia Visual Ontology (~167,000) line of code \[^c\]
- Complete Evaluation Results: \[^d\]
- KNN in Nutch \[^e\]

\[^a\] http://hypermanymedia.wku.edu

\[^b\] http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl

\[^c\] http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml

\[^d\] http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/

\[^e\] http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/
Research Links

Important Links:

- HyperManyMedia: \(^a\)
- HyperManyMedia Ontology (~40,000) line of code \(^b\)
- HyperManyMedia Visual Ontology (~167,000) line of code \(^c\)
- Complete Evaluation Results: \(^d\)
- KNN in Nutch \(^e\)

\(^a\) [http://hypermangimedia.wku.edu](http://hypermangimedia.wku.edu)
\(^b\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl)
\(^c\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml)
\(^d\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/)
\(^e\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/)
Research Links

Important Links:

- HyperManyMedia: \(^a\)
- HyperManyMedia Ontology (~40,000) line of code \(^b\)
- HyperManyMedia Visual Ontology (~167,000) line of code \(^c\)
- Complete Evaluation Results: \(^d\)
- KNN in Nutch \(^e\)

\(^a\) http://hypermanymedia.wku.edu
\(^b\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl
\(^c\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml
\(^d\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/
\(^e\) http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/
Important Links:

- HyperManyMedia: \(^a\)
- HyperManyMedia Ontology (~40,000) line of code \(^b\)
- HyperManyMedia Visual Ontology (~167,000) line of code \(^c\)
- Complete Evaluation Results: \(^d\)
- KNN in Nutch \(^e\)

\(^a\) [http://hypermanymedia.wku.edu](http://hypermanymedia.wku.edu)
\(^b\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/semanticnew.owl)
\(^c\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/VisualOntologyGraph.xml)
\(^d\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/)
\(^e\) [http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/](http://acadmedia.wku.edu/Zhuhadar/Evaluation%20Results/WekaInNutch/)
Linked data–the story so far.

B. Croft, D. Metzler, and T. Strohman.
*Search Engines: Information Retrieval in Practice*.

R. Feldman and J. Sanger.
*The text mining handbook*.

M. Kavouras and M. Kokla.
*Theories of geographic concepts: ontological approaches to semantic integration*.

C.D. Manning, H. Schütze, and MIT Press.
*Foundations of statistical natural language processing*.
O. Nasraoui and E. Saka.

Douglas W. Oard and Bonnie J. Dorr.
A survey of multilingual text retrieval. 1996.

M.A. Sicilia and M.D. Lytras.

Y. Zhao and G. Karypis.