

AgentMatcher Search in Weighted, Tree-Structured Learning Object Metadata

H. Boley, V.C. Bhavsar,
D. Hirtle, A. Singh, Z. Sun and L. Yang
National Research Council and
University of New Brunswick

Learning Objects Summit
Fredericton, NB, Canada, March 29-30, 2004

Outline

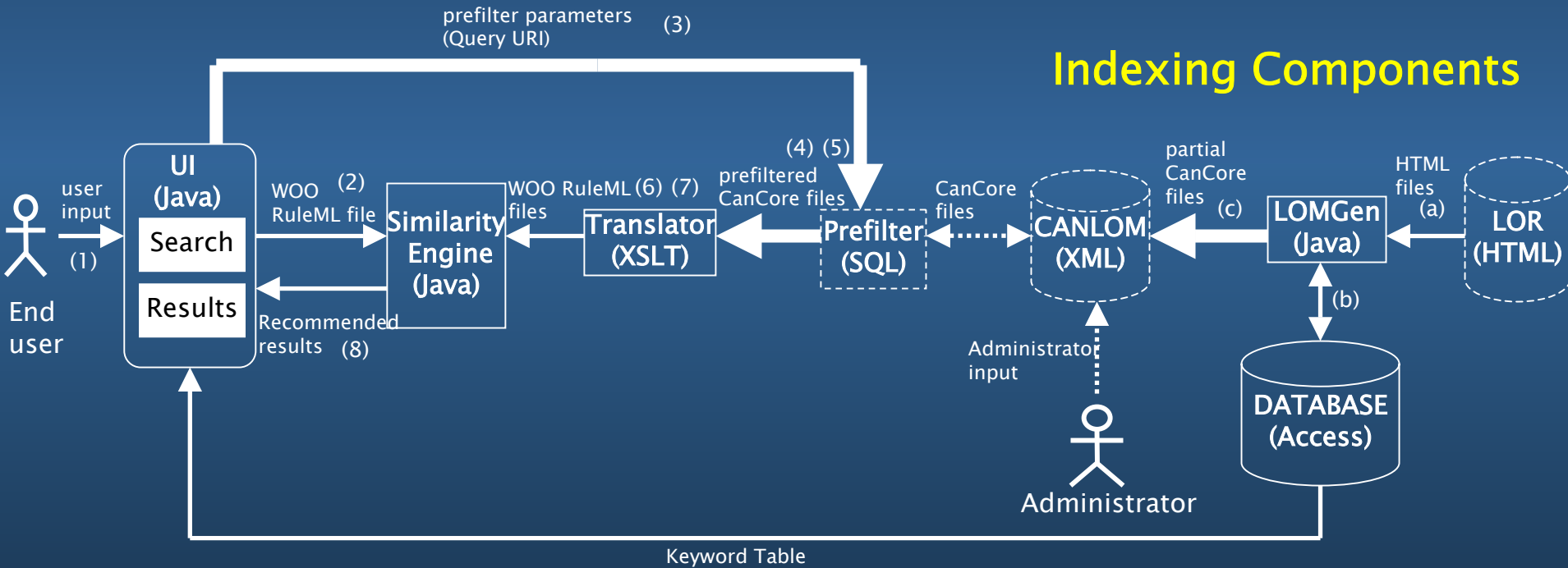
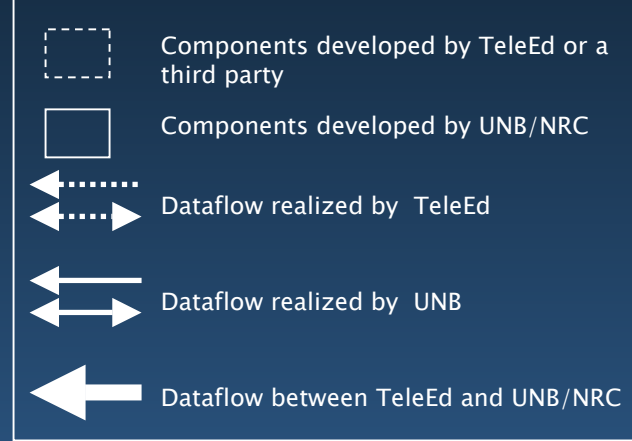
- AgentMatcher Overview
- User Interface
- Translator
- Similarity Engine
- LOM Generator
- Conclusion

AgentMatcher Overview

- AgentMatcher developed for
 - match-making between
 - buyer and seller agents
- Instantiated to
 - learners in search for
 - procurable learning objects (LOs)
described by Learning Object Metadata (LOM)

Architecture

Retrieval Components



Indexing Components

AgentMatcher Components

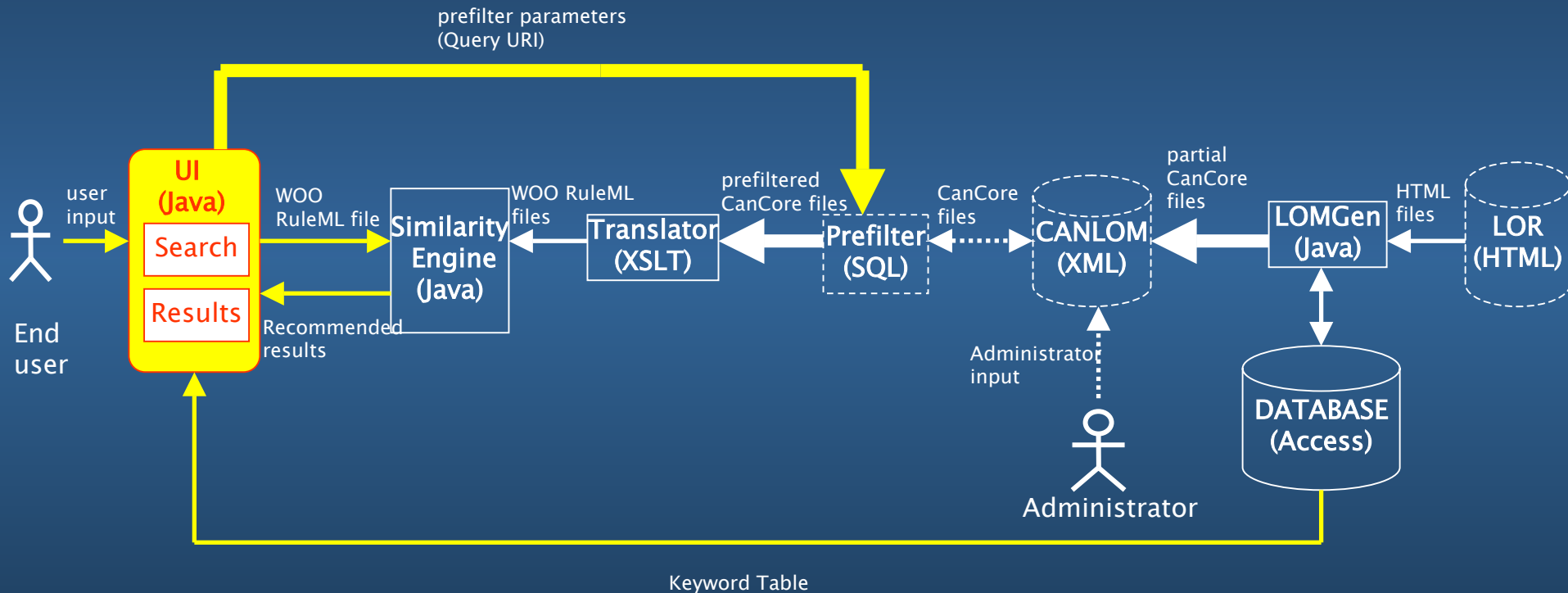
- User Interface
 - Generates prefilter parameters (Query URIs)
 - Generates query trees
 - Communicates between internal and external components
- Translator
 - Transforms LOM from Prefilter to WOO RuleML

AgentMatcher Components (Cont'd)

- Similarity Engine
 - Computes the similarity between query and LOM trees
 - Presents ranked search results as HTML table
- LOMGen (*Learning Object Metadata Generator*)
 - Semi-automatically extracts information from LOs (in *HTML*)
 - Constructs keyword/key phrase database
 - Posts to CANLOM Metadata Repository

User Interface (ZhongWei Sun)

— Overview



- Interacts between user and the advanced search components

User Interface

— Actions

- Retrieves keywords from LOMGen database
- Allows convenient input of tree structured features as well as corresponding weights for them
- Generates a Weighted Object-Oriented RuleML document
- Sends prefiltering query to, and retrieves LOM response stream from, KnowledgeAgora
- Splits the returned response LOM document
- Invokes Translator and Similarity Engine

Overview of User Interface

Advanced Search

Please use the sliders to represent the importance of each choice.

Please hold "ctrl" key to select multiple keywords. Your input is required for keywords.

General

Title:

Keywords:

Language:

Classification

Discipline:

Educational

Intended End User:

Typical Age Range:

Educational Use:

Educational Types:

Typical End User Language:

Lifecycle

Contributor(Publisher):

Price:

Technical

Format:

Technical Platform Requirements:

Rights

Copy Right Restrictions:

Copy Right Description:

Please input a threshold for the result display:

- Features within the same CanCore schema category are grouped within one box
- “Don’t care” is in green to tell the user it will match arbitrary values
- Default values except for “Don’t care” are in red to remind the user of potential need for change
- The relative importance of each feature is entered via sliders from 0.0 to 1.0, summing up to 1.0⁸

User Interaction

- User inputs tree structured features and assigns a weight to each of them
 - This weight indicates the feature's importance to the user relative to other features within the same category
 - Weights can be entered with sliders, where all other weights of that group self-adjust to keep their sum at 1.0
- User inputs a threshold as a cut-off line for the final result recommendation
 - The learning objects with similarities above the threshold are considered as recommended

User Interface (Screenshot 1)

Advanced Search

Please use the sliders to represent the importance of each choice.

Please hold "ctrl" key to select multiple keywords. Your input is required for keywords.

General

Title: Introduction to Oracle (Slider: 0.3)

Keywords: neural network, operating system, Oracle, Oracle Corporation (Slider: 0.5)

Language: en (Slider: 0.2)

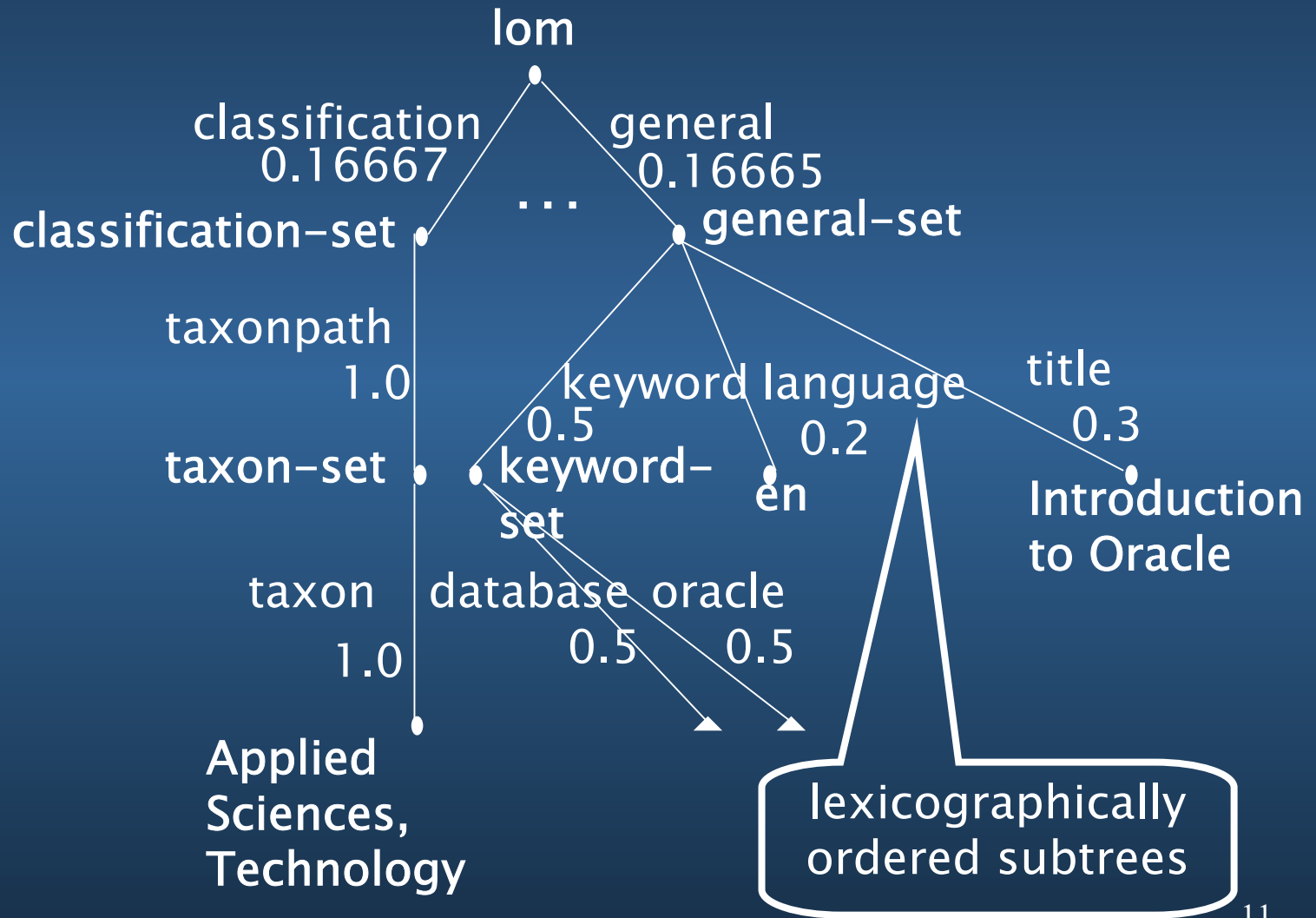
Classification

Discipline: Applied Sciences, Technology (Slider: 1.0)

"database"
(not visible
here) and
"Oracle"
are chosen as
Keywords

cannot be
changed,
hence
grayed out

Corresponding Query Tree (1)



User Interface (Screenshot 2)

The screenshot shows a software window titled "Advanced Search" with a blue header and standard window controls. The interface is divided into two main sections: "Educational" and "Lifecycle".

Educational Section:

- Intended End User:** A dropdown menu with "Learner" selected.
- Typical Age Range:** A dropdown menu with "18-25" selected.
- Educational Context:** A dropdown menu with "Don't care" selected.
- Educational Types:** A dropdown menu with "Course" selected.
- Typical End User Language:** A dropdown menu with "Don't care" selected.

Each dropdown menu in the Educational section is accompanied by a horizontal slider control. The sliders are positioned as follows:

- Intended End User: Slider at 0.3 (range 0 to 1.0)
- Typical Age Range: Slider at 0.1 (range 0 to 1.0)
- Educational Context: Slider at 0.0 (range 0.0 to 1.0)
- Educational Types: Slider at 0.6 (range 0 to 1.0)
- Typical End User Language: Slider at 0.0 (range 0.0 to 1.0)

Lifecycle Section:

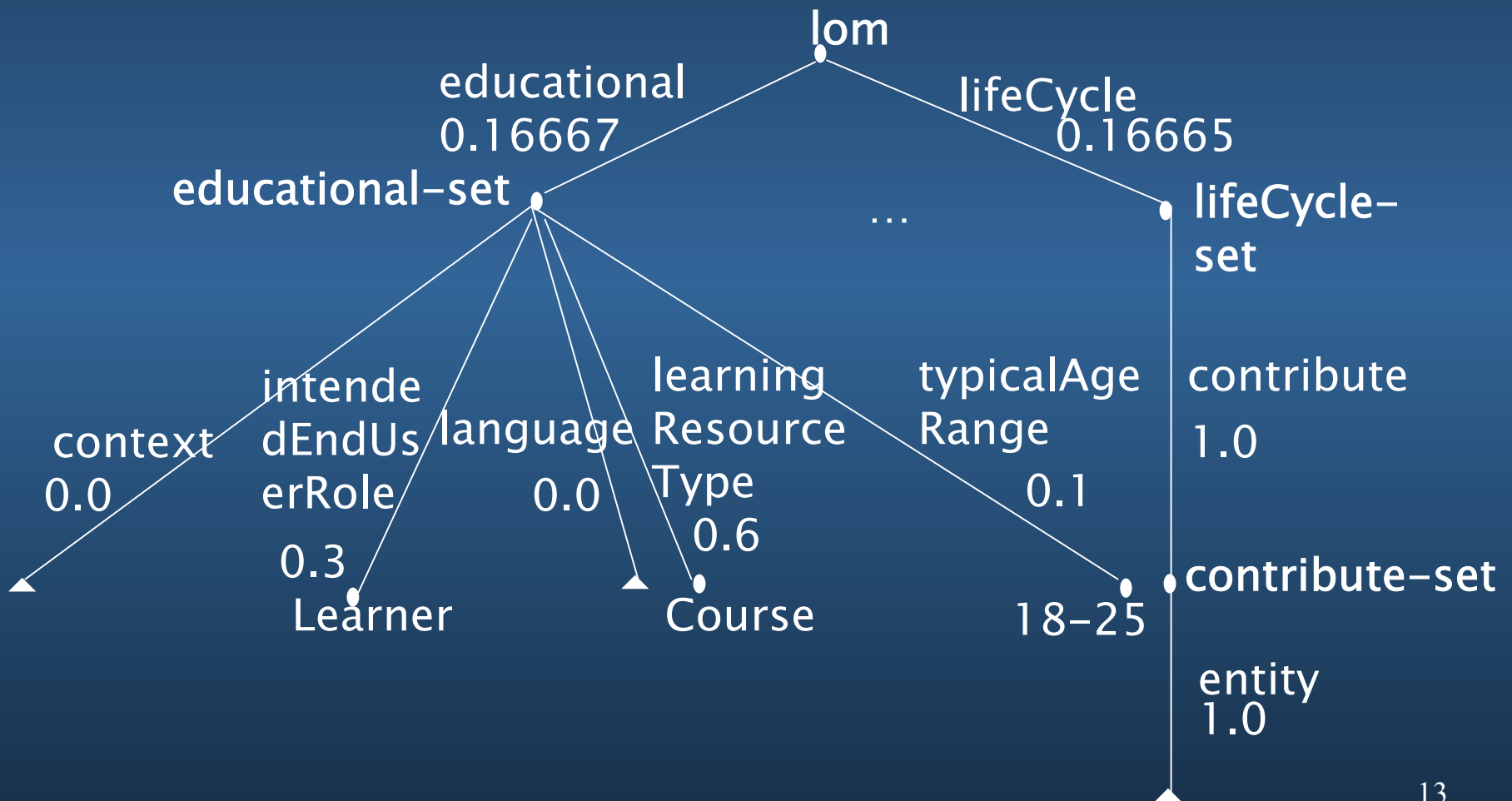
- Contributor (Publisher):** An empty text input field.
- Price:** An empty text input field.

Each input field in the Lifecycle section is accompanied by a horizontal slider control, both positioned at 0 (range 0 to 1.0).

kept empty since not standardized and not important to students

non-CanCore element, hence not filled out

Corresponding Query Tree (2)



User Interface (Screenshot 3)

The screenshot shows a window titled "Advanced Search" with a blue title bar and standard window controls. The interface is divided into three main sections: "Technical", "Rights", and "Threshold for result cut-off".

Technical Section:

- Format:** A dropdown menu set to "HTML". To its right is a slider control ranging from 0 to 1.0, with the current value set to 0.6.
- Technical Platform Requirements:** A dropdown menu set to "Don't care". To its right is a slider control ranging from 0 to 1.0, with the current value set to 0.4.

Rights Section:

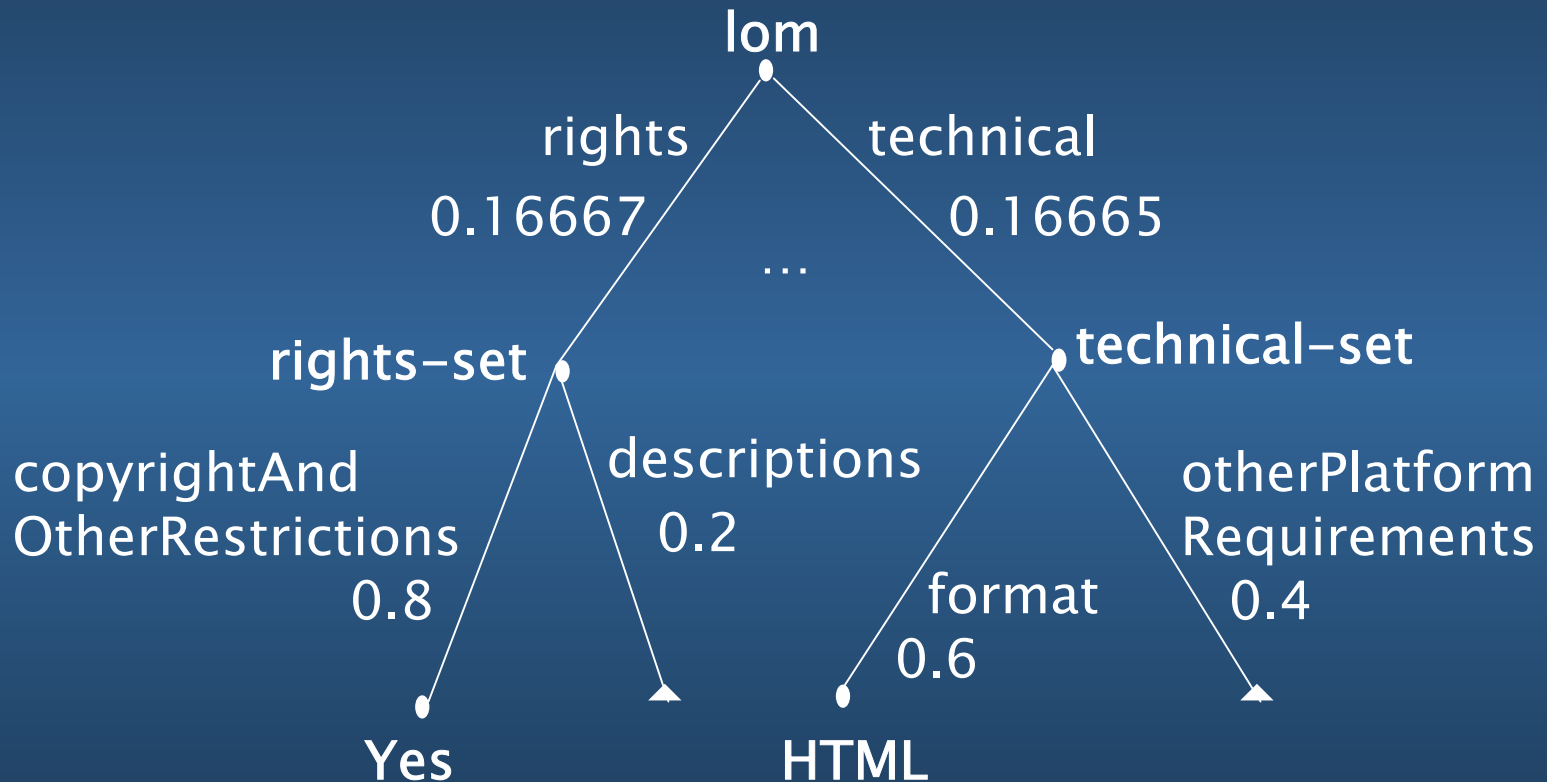
- Copy Right Restrictions:** A dropdown menu set to "Yes". To its right is a slider control ranging from 0 to 1.0, with the current value set to 0.8.
- Copy Right Description:** A dropdown menu set to "Don't care". To its right is a slider control ranging from 0 to 1.0, with the current value set to 0.2.

Threshold for result cut-off: A slider control ranging from 0 to 1.0, with the current value set to 0.7.

At the bottom of the window, there are two buttons: "Submit" and "Clear".

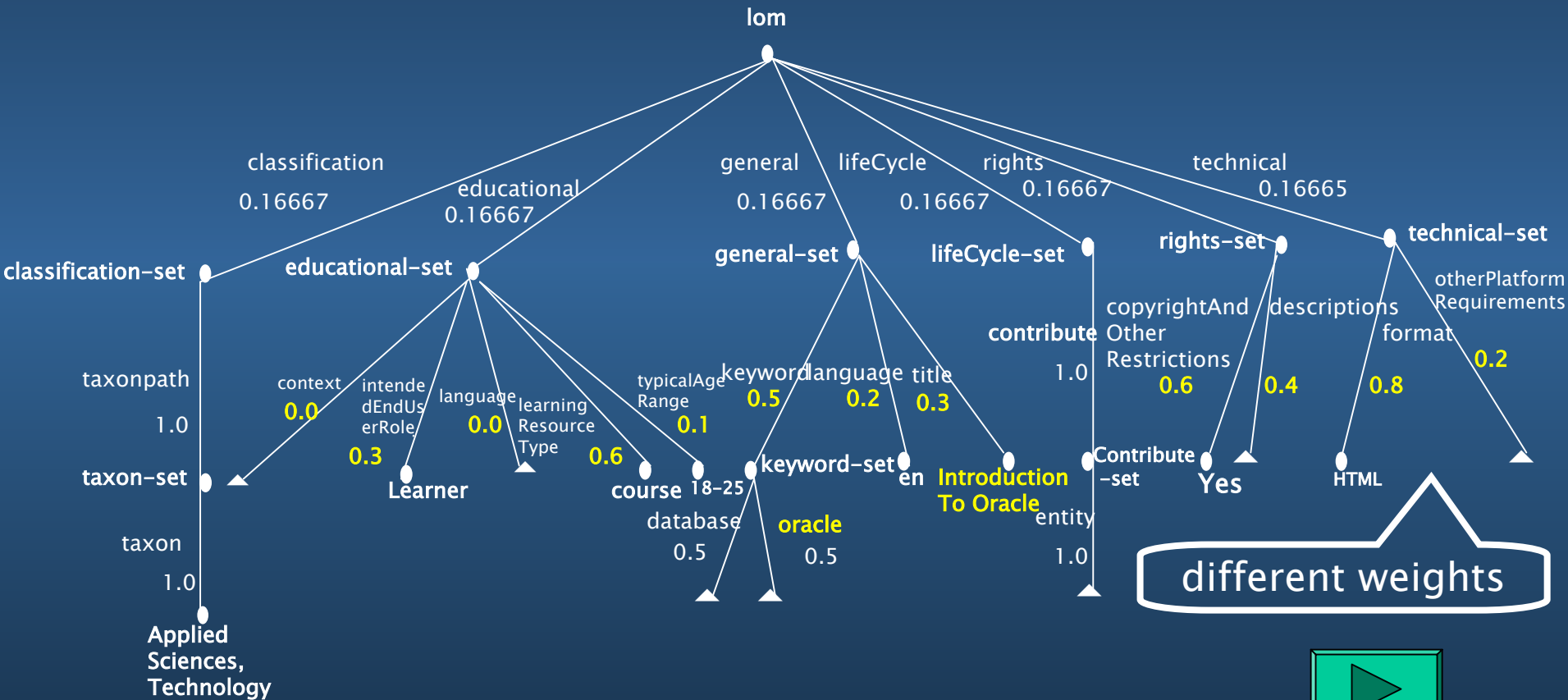
Threshold

Corresponding Query Tree (3)



Overall Query Tree

—From User Interface



User Interface Outputs

- Query URI sent to Prefilter:

http://www.knowledgeagora.com/urb_search/search_results.cfm?requestTimeout=300&query=database+oracle&selectedLanguageID=1&selectedgranID=1&maxPrice=&UpperCategory=184&publisherQuery=&selectedEduContext=&selectedCopyRightValue=Yes&selectedSoftwareRequirements=&selectedTechFormat=HTML

encodes "en"

encodes
"course"

encodes
"Applied Science,
Technology"

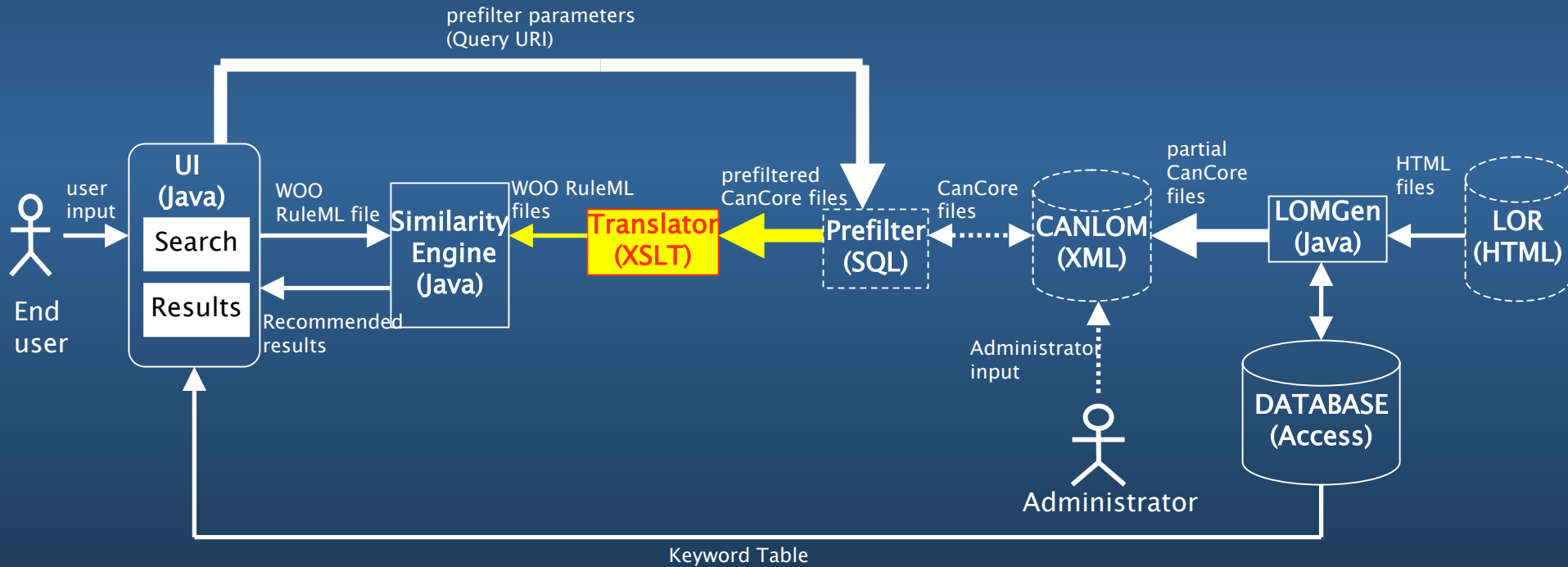
- XML document sent to Similarity Engine
 - both tree structure and branch weights
- Threshold sent to Similarity Engine
(0.7 is sent in this example)

User Interface — Summary

- Enables convenient interaction between user and the internal components
- Communicates between the internal components and the external KnowledgeAgora website
- Invokes both internal and external components

Translator (David Hirtle)

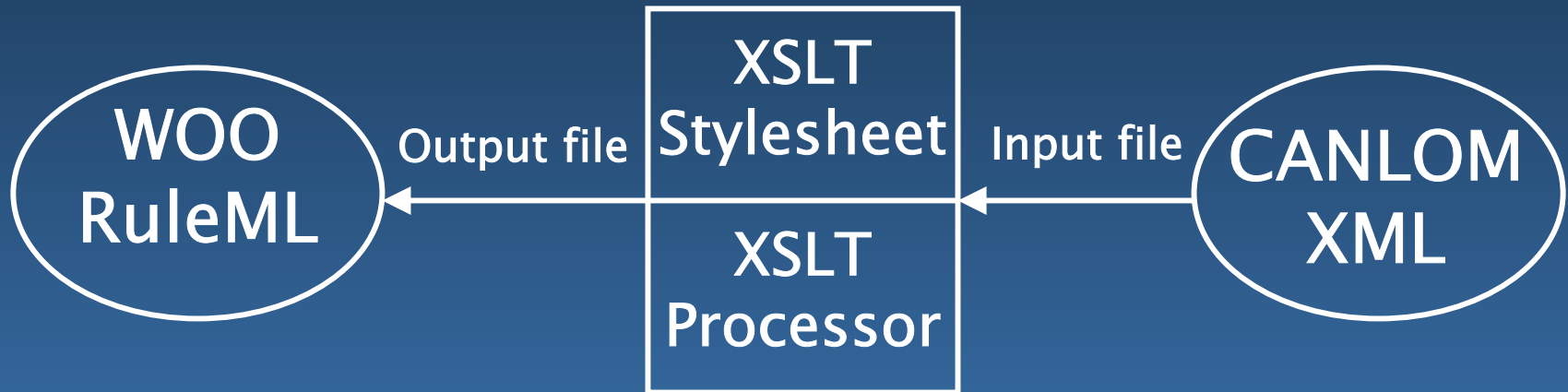
Overview



Translator Overview

← - - Similarity Engine

Prefilter - - →



- Input: CANLOM XML (from Prefilter)
- Output: Weighted Object-Oriented RuleML (to Similarity Engine)
- Tool: Extensible Stylesheet Language Transformations (XSLT)
 - Language for transforming XML to XML (W3C Rec)

Sample Translation

WOO RuleML

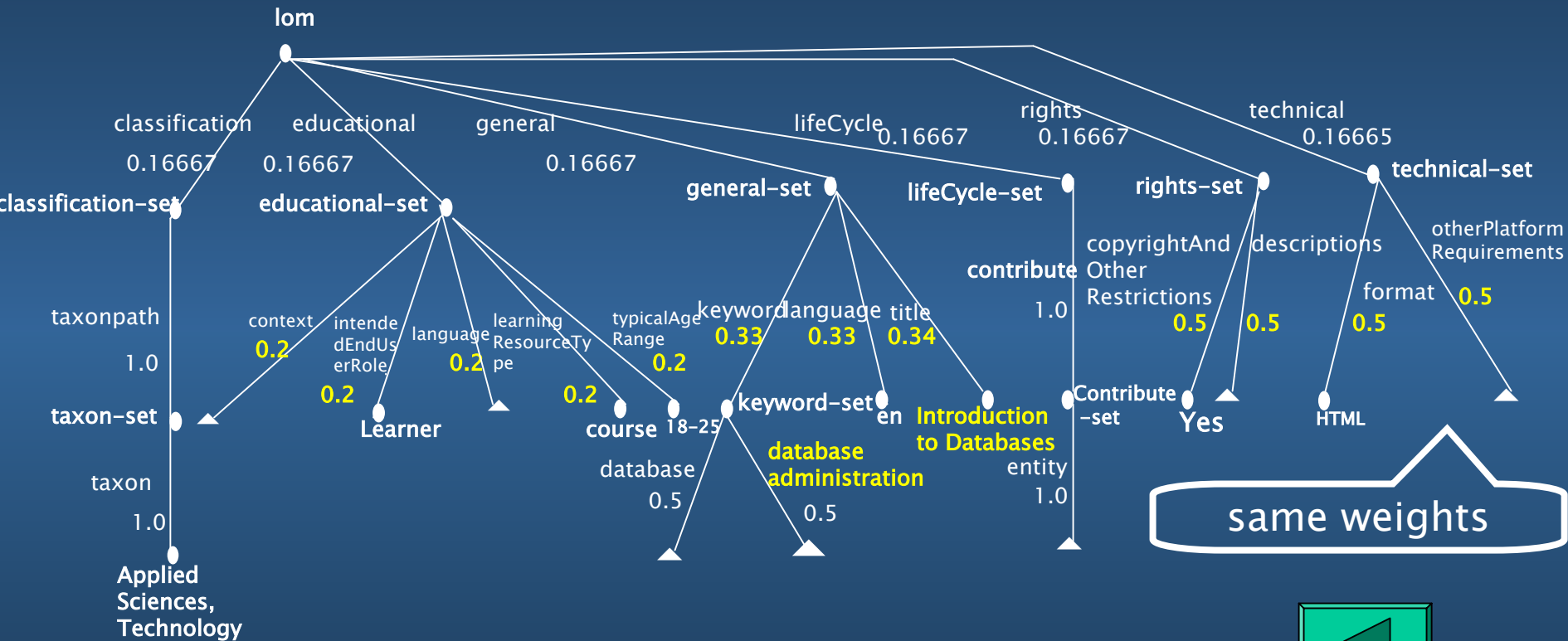


CANLOM XML

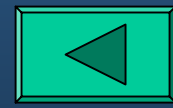
```
<cterm>
  <_opc>
    <ctor>lom</ctor>
  </_opc>
  ...
  <_slot name="general" weight="0.16667">
    <cterm>
      <_opc>
        <ctor>general_set</ctor>
      </_opc>
      ...
      <_slot name="title" weight="0.33333">
        <ind>
          Introduction to Databases
        </ind>
      </_slot>
      ...
    </cterm>
  </_slot>
  ...
</cterm>
```

```
<lom>
  <general>
    ...
    <title>
      <string>
        Introduction to
        Databases
      </string>
    </title>
    ...
  </general>
  ...
</lom>
```

Translated Weighted LOM Tree



same weights



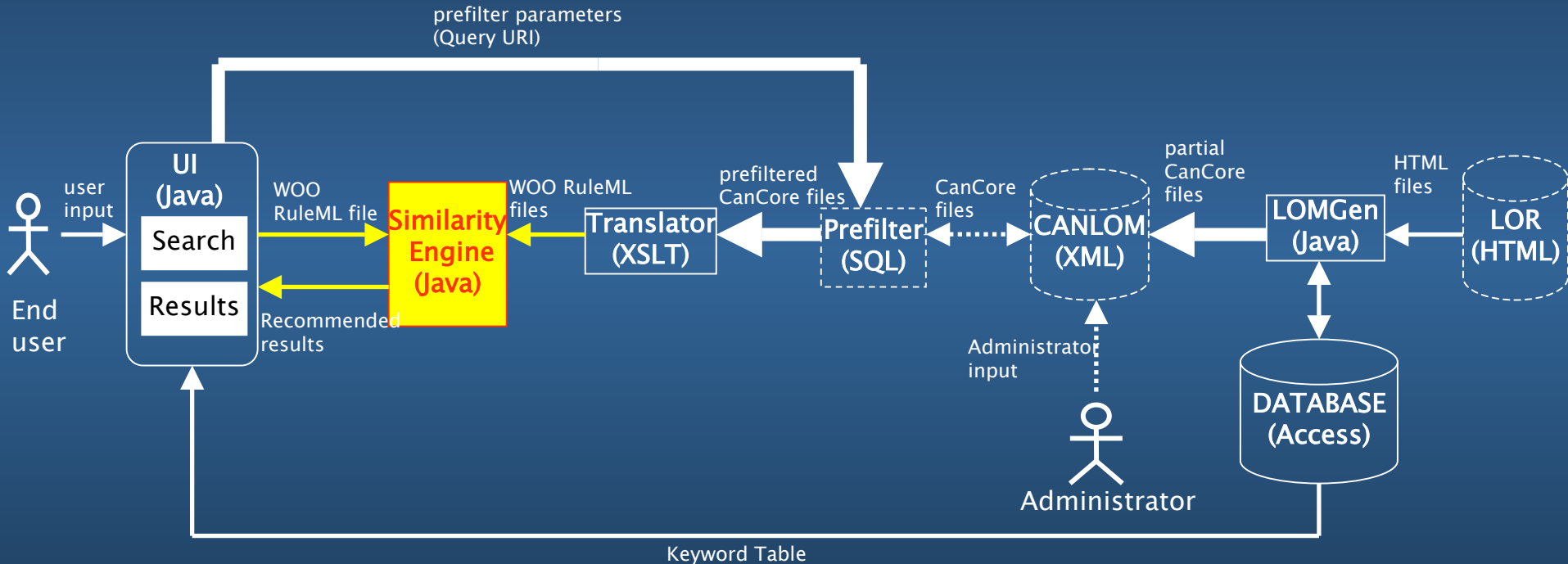
- Different weights possible in LOM Tree, like in Query Tree

Translator Summary

- Uses XSLT engine embedded in main program
- Transforms CANLOM XML returned from Prefilter into WOO RuleML
- Resulting WOO RuleML then used by Similarity Engine

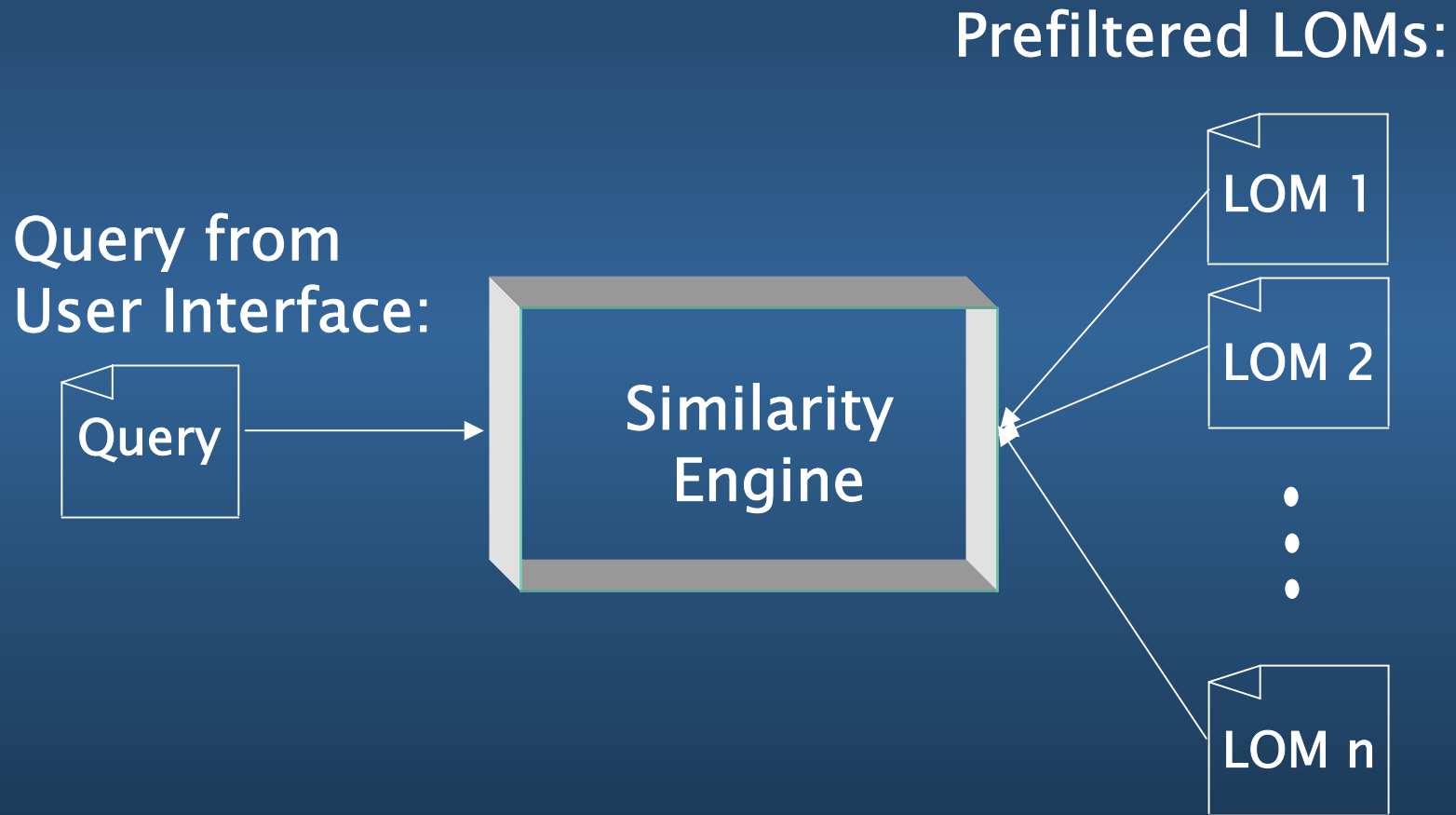
Similarity Engine (Lu Yang)

— Overview





- Computes similarity between query and LOM trees
- Displays ranking search results in a browser

Similarity Engine — Inputs



Similarity Engine

— Tree Similarity Algorithm

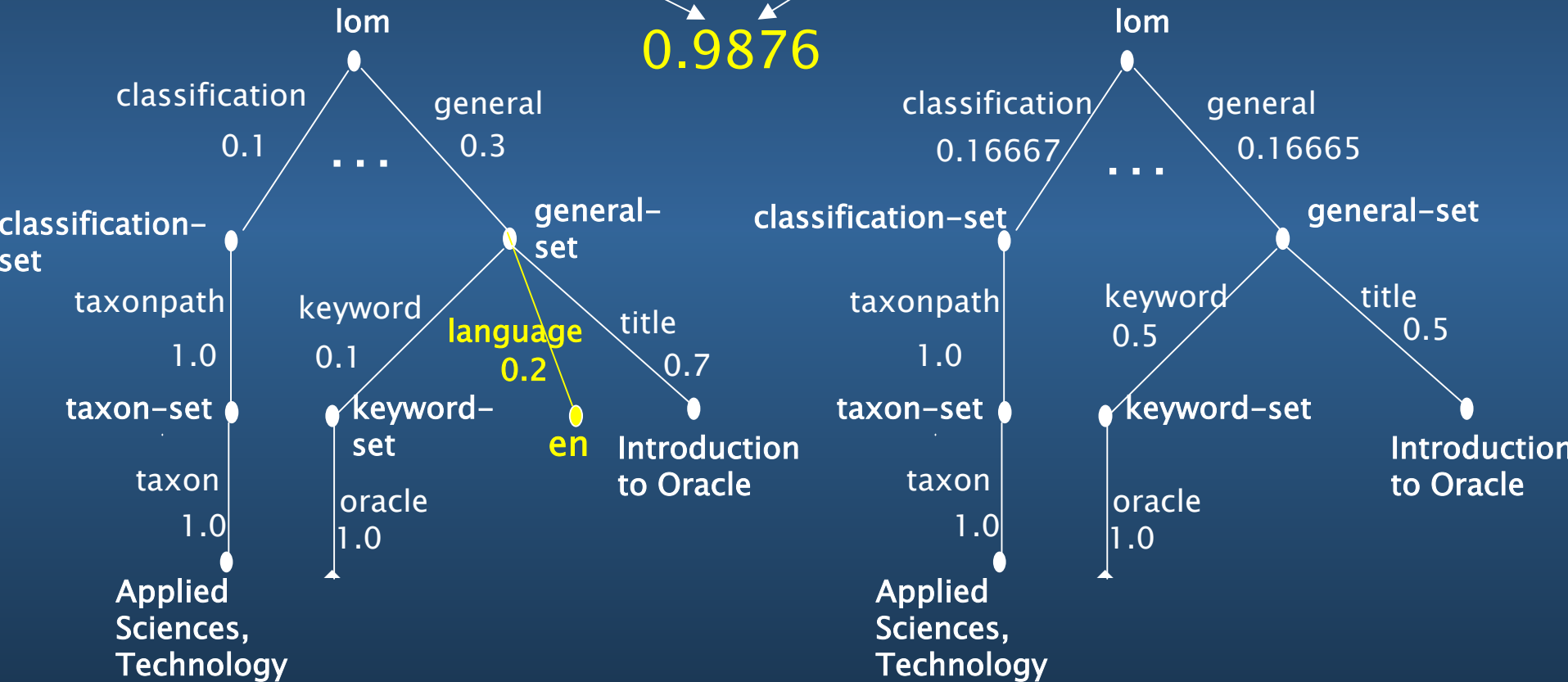
- Tree representations 
 - Arcs at the same level of a (sub)tree are in lexicographic order
 - For linear left-to-right traversal
 - Arc weights at the same level of a (sub)tree add up to 1.0
 - No limit on tree breadth and depth
 - Don't care nodes allowed
- Tree similarity algorithm [1,2] 
 - Similarity values fall into [0.0,1.0]
 - Top-down tree traversal
 - Bottom-up value propagation

Similarity Engine

Similarity Computation (example)

A Query Tree (user.xml)

A LOM Tree (WOORuleML2.xml)



Similarity Engine

— Results (low threshold)

Search Results - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 搜索 收藏夹 媒体 打印 打印 打印 打印

地址(A) E:\SimilarityForKnowledgeAgora\onHia3\index.html 转到 链接

Here are your search results with threshold 0.86;
click LOM or LO links to get details.

Rank	Similarity	LOMs	LOs
1	0.9876	WOORuleML2.xml	Go to the website
2	0.9769	WOORuleML10.xml	Go to the website
3	0.9340	WOORuleML6.xml	Go to the website
4	0.9123	WOORuleML12.xml	Go to the website
5	0.9008	WOORuleML11.xml	Go to the website
6	0.8762	WOORuleML8.xml	Go to the website
7	0.8713	WOORuleML9.xml	Go to the website
8	0.8647	WOORuleML1.xml	Go to the website
0.86 (threshold)			
9	0.8014	WOORuleML4.xml	Go to the website
10	0.7435	WOORuleML7.xml	Go to the website
11	0.5032	WOORuleML3.xml	Go to the website
12	0.4109	WOORuleML5.xml	Go to the website

Similarity Engine

— Results (high threshold)

Search Results - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 刷新 地址 搜索 收藏夹 媒体 打印 打印范围 打印内容 打印页眉 打印页脚 打印范围 打印内容 打印页眉 打印页脚

地址 @ E:\SimilarityForKnowledgeAgoraYouHua3\noResults.html 转到 链接

Sorry, there is no matching LO having similarity above your threshold: 0.99.

Rank	Similarity	LOMs	LOs
0.99 (threshold)			
1	0.9876	WOORuleML2.xml	Go to the website
2	0.9769	WOORuleML10.xml	Go to the website
3	0.9340	WOORuleML6.xml	Go to the website
4	0.9123	WOORuleML12.xml	Go to the website
5	0.9008	WOORuleML11.xml	Go to the website
6	0.8762	WOORuleML8.xml	Go to the website
7	0.8713	WOORuleML9.xml	Go to the website
8	0.8647	WOORuleML1.xml	Go to the website
9	0.8014	WOORuleML4.xml	Go to the website
10	0.7435	WOORuleML7.xml	Go to the website
11	0.5032	WOORuleML3.xml	Go to the website
12	0.4109	WOORuleML5.xml	Go to the website

Similarity Engine — Summary

- Functionality of Similarity Engine:
 - Similarity computation
 - Ranked result display
- The core of Similarity Engine
 - Tree similarity algorithm
- Threshold for cut-off
- Ranking and HTML output of LOMs and LOs

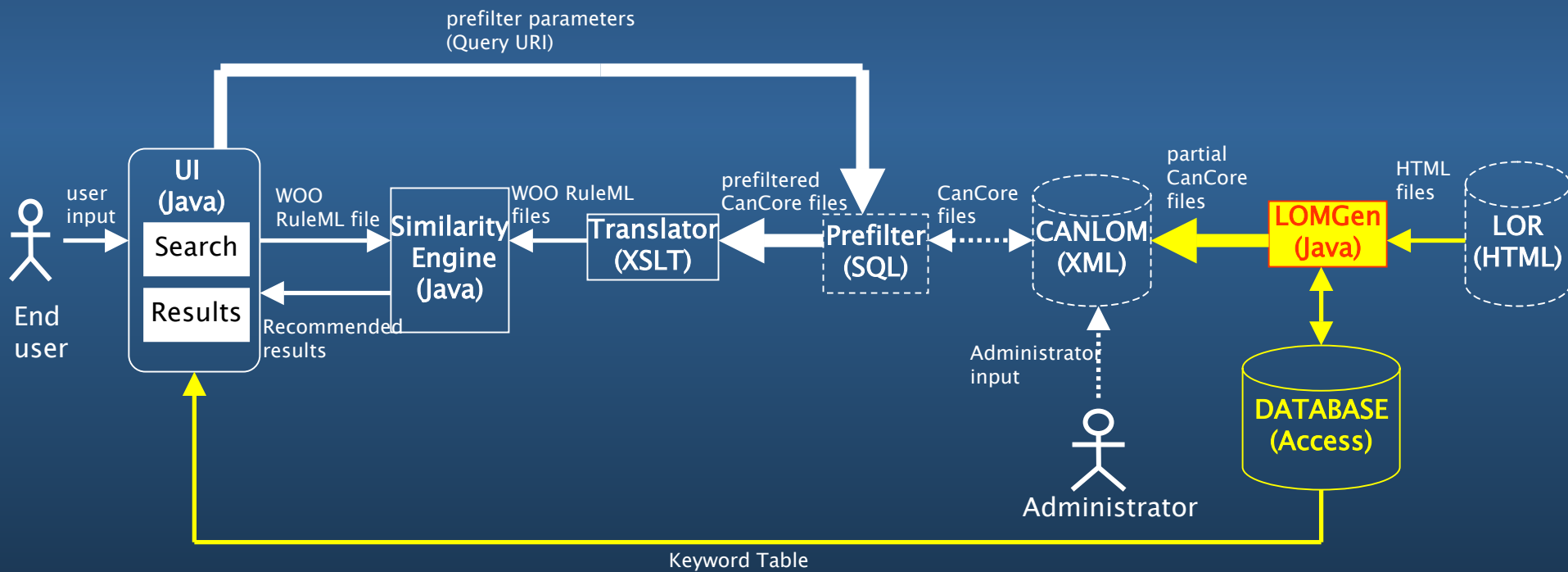
Similarity Engine — Reference

[1] Bhavsar, V.C., H. Boley, L. Yang, “A Weighted-Tree Similarity Algorithm for Multi-Agent Systems in E-Business Environments”, *In* Proceedings of 2003 Workshop on Business Agents and the Semantic Web, Halifax, June 14, 2003, National Research Council of Canada, Institute for Information Technology, Fredericton, pp. 53-72, 2003. To appear *in* Computational Intelligence, November 2004.



LOM Generation (Anurag Singh)

LOMGen Overview



LOMGen

——— Functionality

- Extracts keywords and keyphrases from an HTML document
- Derives a dictionary of “surface terms” and “deep terms” from Free Online Dictionary of Computing (FOLDOC)
- Provides a GUI for metadata administrator
- Makes up-to-date list of keyphrases and keywords available to the AgentMatcher UI
- Makes similarity computation more accurate
- Automates metadata (XML) posting to LOM repository

LOMGen

Administrator GUI

- The keywords and keyphrases extracted from the HTML file are presented to the metadata administrator
- The administrator can select and add any additional terms they find are important
- The administrator is also prompted to select a category for the LO

The screenshot shows the LOMGen Learning Object Metadata Generator Administrator GUI. The window title is "LOMGen Learning Object Metadata Generator". The main area is divided into two columns. The left column, titled "KEYPHRASE", contains a list of checkboxes for various terms: OWL, semantic web, RDF, web site, Internet, World-Wide Web, resource description framework, W3C, meta-data, computer science, informatics, web, Donald Knuth, and literate programming. The right column, titled "ADD SYNONYMS (optional)", contains a list of text boxes, with the first one containing "Web Ontology Language". Below the "KEYPHRASE" list is a text box labeled "Add more terms to describe the Learning Object (Optional)". At the bottom left, there are "Delete" and "Add" buttons. At the bottom right, there is a dropdown menu labeled "Select a domain term to describe the Learning Object" with "World-Wide Web" selected. An "OK" button is located at the bottom center.

LOMGen

—— Summary

- LOs analyzed for LOM content:
 - Meta tags **title**, **description**, **keywords** of HTML **head** turned out to be directly usable for dictionary
 - Shallow parsing of natural language HTML **body** further processed interactively with the user
- Dual use of LOMGen:
 - AgentMatcher UI keywords
 - KnowledgeAgora metadata repository

AgentMatcher Conclusion

- UI to input LO search features structured as LOM-like trees with relative branch weights
- LOs prefiltered from CANLOM repository, then translated for use by Similarity Engine
- Similarity Engine computes similarity between query and prefiltered results using weighted tree similarity algorithm
- LOs ranked and displayed to user in HTML
- LOMGen generates metadata from a Learning Object Repository and posts them to the CANLOM database
- Demo's next, as time permits