Design and Implementation of an Academic Search System Based on a General Query Language and Automatic Question Answering

Zi Xiong, Yue Qi, Qikai Cheng, Wei Lu
INTRODUCTION

• Design of a general query language for academic search.

• Understanding the search intentions of users’ questions.

• Implemented system of QA-oriented academic search engine.
SSL: SCHOLAR SPECIFIC LANGUAGE

- An expansion of traditional DSL.

3 main modules:
1. **Type module**: Represent the type of purpose information or intention.
2. **Field module**: Which expresses the query mode which is the specific combination of destination information attributes.
3. **Refinement module**: Which is used to represent the refined query semantics of the result information, includes both the post-filtering semantics and the secondary retrieval semantics.
SSL: SCHOLAR SPECIFIC LANGUAGE

- intent = "" module ""
- module = type "," field [",", refinement]
- type = DQUOTE "type" DQUOTE ":" intent-category
- field = DQUOTE "field" DQUOTE ":" fields
- refinement = DQUOTE "refinement" DQUOTE ":" refinements
- intent-category = DQUOTE ("paper"/"citation"/"entity"/"concept"/"qa") DQUOTE
- item = DQUOTE [bool] 1*char DQUOTE
- item = DQUOTE [bool] 1*char DQUOTE
- bool = "+" / "]" / "-
- refinements = "" quantity "," rank ["," field] ""
- quantity = DQUOTE "quantity" DQUOTE ":" quantity-item
- quantity-item = ":" / number
- rank = DQUOTE "rank" DQUOTE ":" rank-item
- rank-item = DQUOTE ("relevance" / "citations" / "download_num" / "time") DQUOTE
- number = D "DIGIT"
- D = "1" / "2" / "3" / "4" / "5" / "6" / "7" / "8" / "9"
- char = unescaped / escaped
- unescaped = %x20-21 / %x23-5B / %x5D-FF
- escaped = %x5C ("" / \\" / \\"b" / \\"f" / \\"n" / \\"r" / \\"t" / (\"u4(ALPHA / DIGIT)))
UNDERSTANDING USER INTENTION

• Intention Recognition
Use DL-based text classifiers to infer user’s search intention through dialog text (MV-LSTM, 2016).
UNDERSTANDING USER INTENTION

Experiment:

- **Dataset** Baidu Academic and Luojia Academic search
- Labelled with 4 labels:
  - 0 literature query
  - 1 academic entities query
  - 3 academic concept query
  - 4 free question-answering

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision\textsubscript{macro}</th>
<th>Recall\textsubscript{macro}</th>
<th>F1\textsubscript{macro}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>0.6969</td>
<td>0.6952</td>
<td>0.6855</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.9506</td>
<td>0.9549</td>
<td>0.9505</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9136</td>
<td>0.9124</td>
<td>0.9126</td>
</tr>
<tr>
<td>MV-LSTM(Ours)</td>
<td><strong>0.9519</strong></td>
<td><strong>0.9512</strong></td>
<td><strong>0.9513</strong></td>
</tr>
</tbody>
</table>
# QA-ORIENTED ACADEMIC SEARCH ENGINE

![Diagram of QA-Oriented Academic Search Engine]

## Application Layer
- Interactive Question and Answer
- Academic search

## Technique Layer
### Question Answering Module
- Speech Recognition
- Intention Recognition
- Dialog Management
- Question Parsing
- Speech Synthesis

### Retrieval Module
- Search Language Parsing
- Information Retrieval
- Wiki Data
- Literature Data
- Indexing Data
- Vocabulary Resources
- Search Log Data

## Data Layer
- Server
- Storage
- Network
- Backup
Thanks For Listening