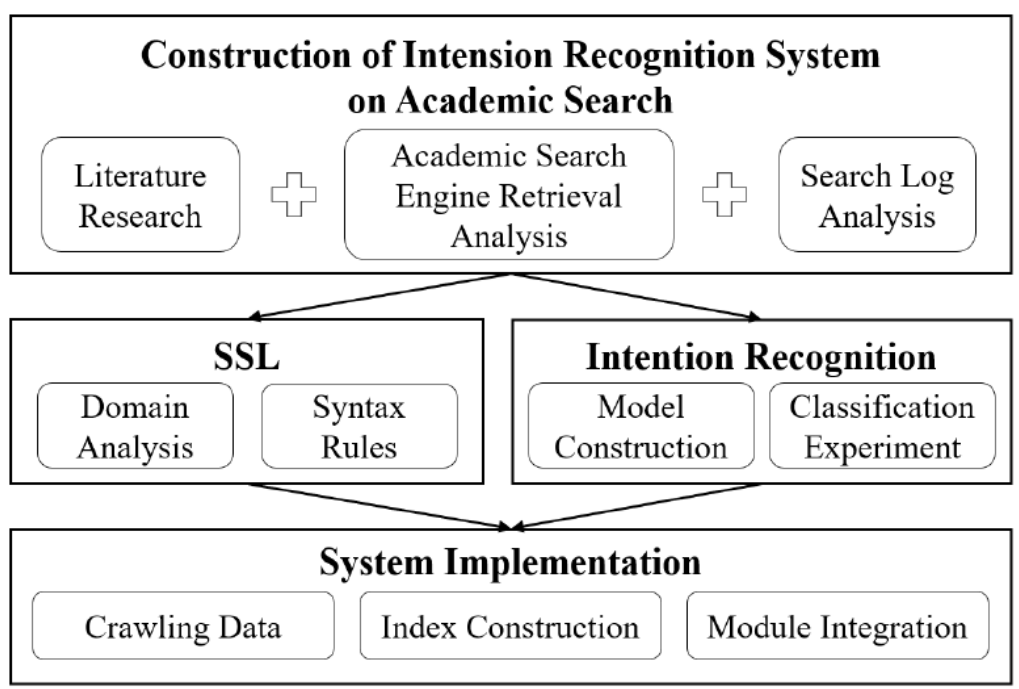
Design and Implementation of an Academic Search System Based on a General Query Language and Automatic Question Answering

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.

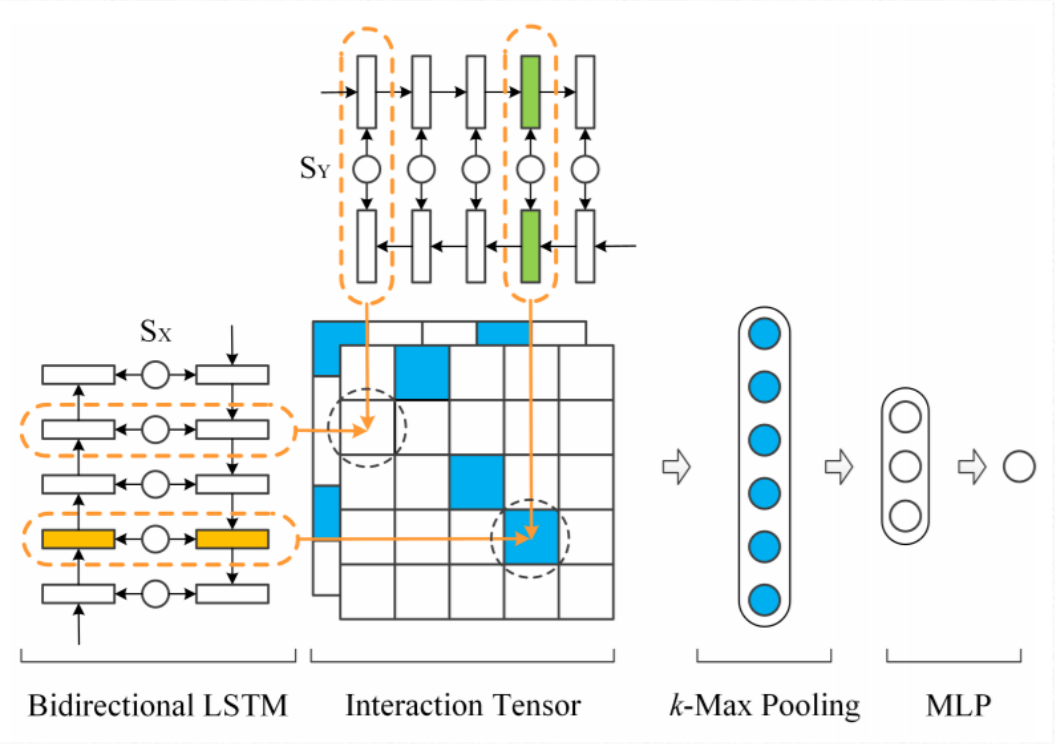
SSL: SCHOLAR SPECIFIC LANGUAGE

- 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
- name = "subject", "author", "time", "title", "keywords", "abstract", "source", "institution", "foundation", "doi", "classification_code", "content", "paper_type", "journal", "conference", "question", "concept"
- 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 = "-1" / number
- rank = DQUOTE "rank" DQUOTE ":" rank-item
- rank-item = DQUOTE ("relevence" / "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" / ("u"4(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

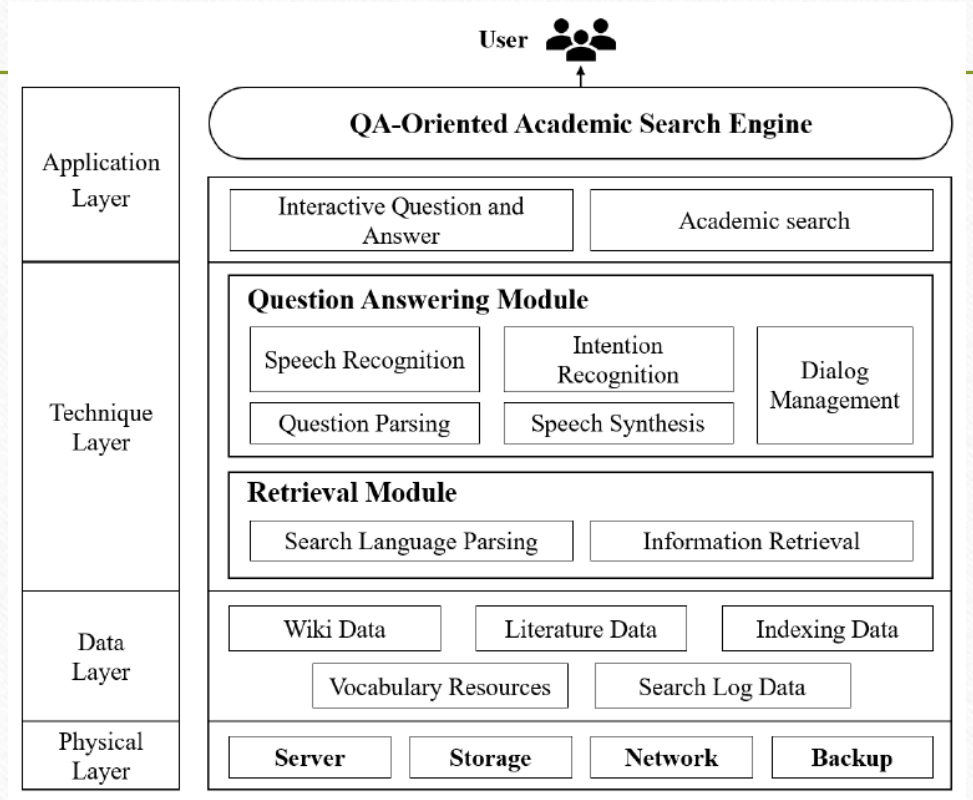
1 academic entities query

3 academic concept query

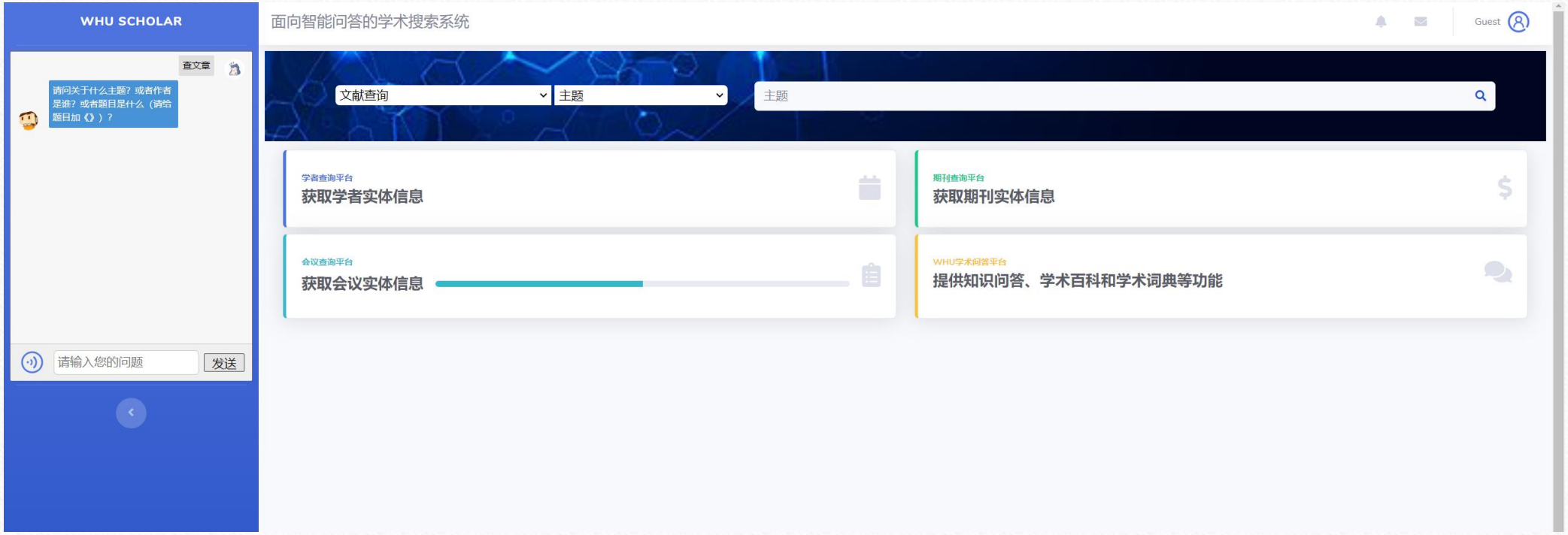
4 free question-answering

Methods	Precision_{macro}	Recall_{macro}	F1_{macro}
Naive Bayes	0.6969	0.6952	0.6855
Logistic Regression	0.9506	0.9549	0.9505
SVM	0.9136	0.9124	0.9126
MV-LSTM(Ours)	**0.9519**	**0.9512**	**0.9513**

QA-ORIENTED ACADEMIC SEARCH ENGINE



System Demo


WHU SCHOLAR
查文章
请问关于什么主题？或者作者是谁？或者题目是什么（请给题目加《》）？
请输入您的问题
发送
面向智能问答的学术搜索系统
Guest
文献查询
主题
主题
学者查询平台
获取学者实体信息
期刊查询平台
获取期刊实体信息
会议查询平台
获取会议实体信息
WHU学术问答平台
提供知识问答、学术百科和学术词典等功能

Thanks For Listening