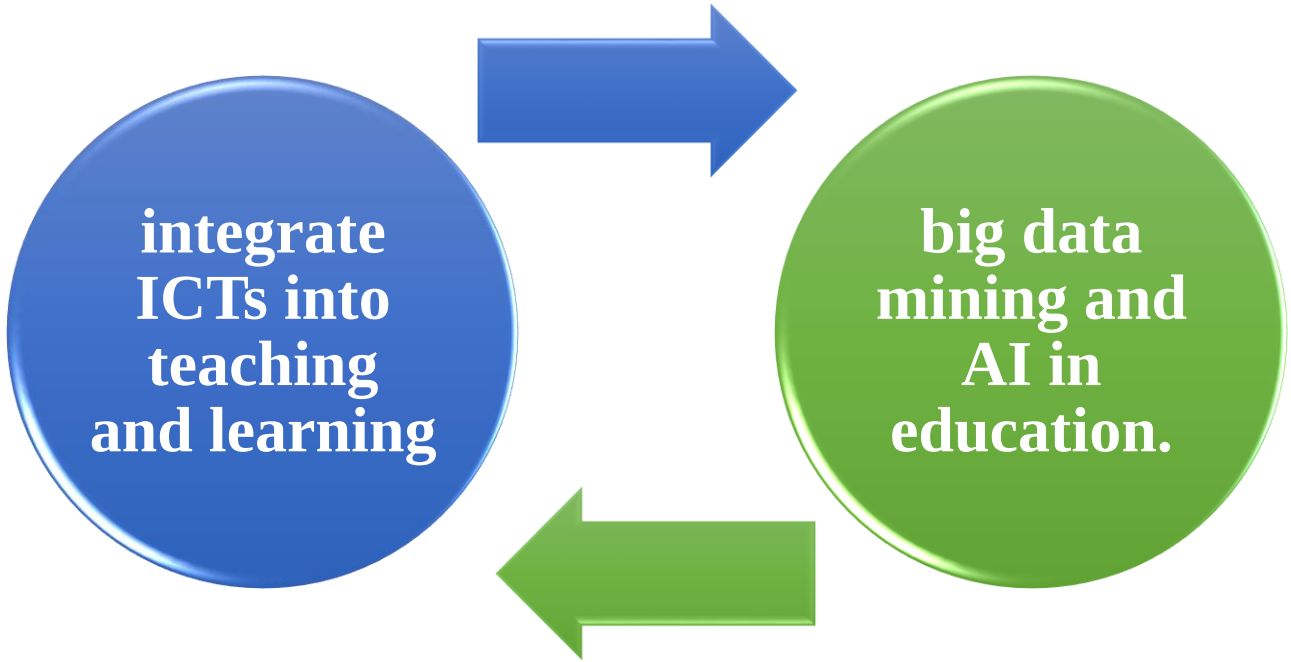
Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

An introduction to the AI tutor project: several ongoing research on big data and artificial intelligence in education

Dr. Baoping Li

Introduction of ICT Center in China

ICT Center of China focuses on research and practice



A learning platform named Smart Learning Partner was developed to support the research and practice

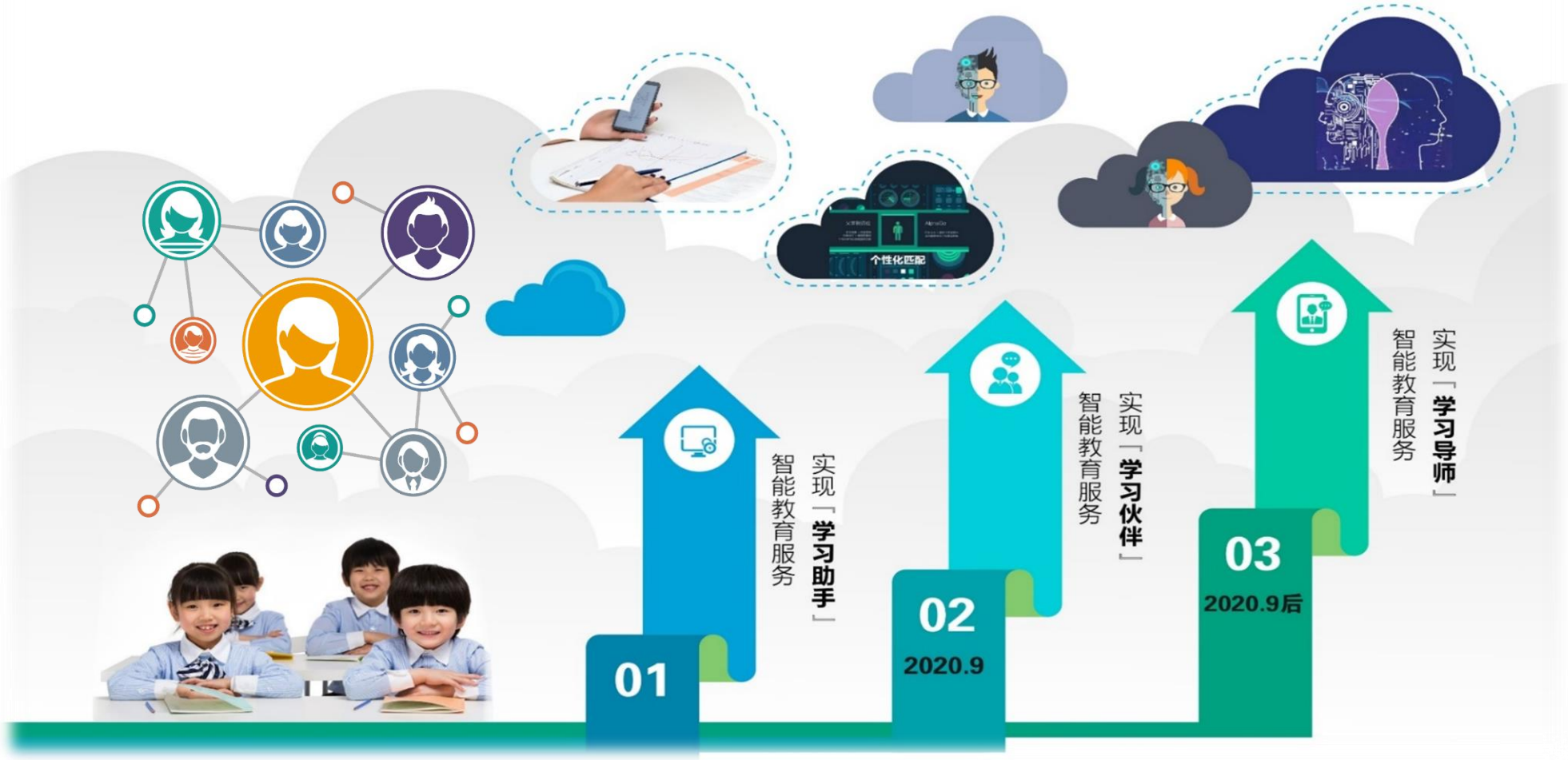
Sino-Finnish Joint Learning Innovation Institute 中芬联合学习创新研究院

SMART LEARNING PARTNER

‘AI Tutor’ Project

Vision: Learning Assistant first, then Learning Partner and AI Tutor finally

To build a comprehensive simulation of the knowledge, emotion, cognition and social network of young children and teenagers so as to provide "intelligent tutor" service with natural language interaction through collecting data and understanding the general rules and individual characteristics of the development of young people.



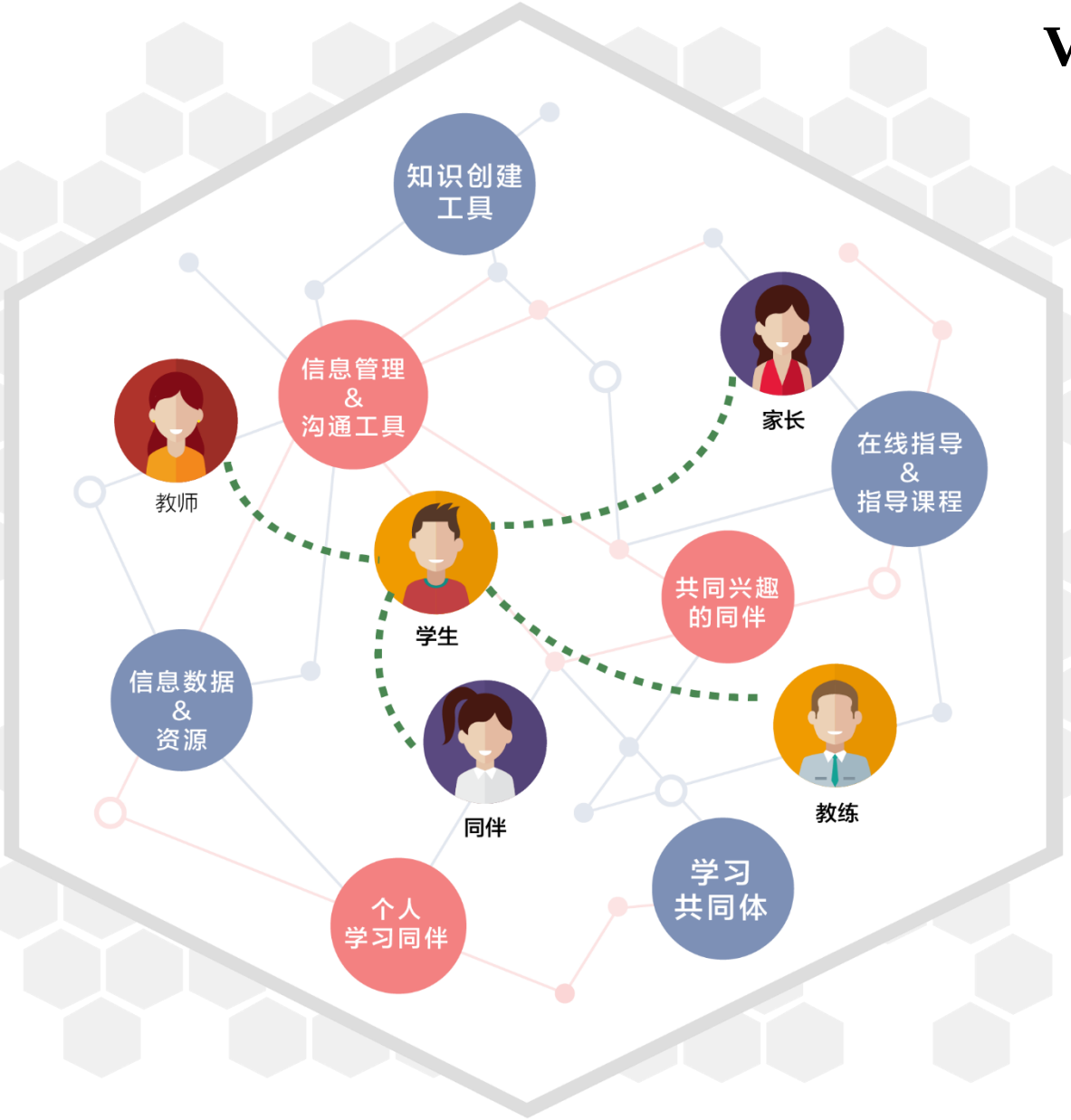
Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Vision: to found a future school with organization innovation

- To found an Internet + supported future school to explore organizational innovation;
- Integration of AI tutor's online teaching service and offline teaching to achieve personalized education;
- To promote project-based exploratory learning cross disciplines and develop students' innovative spirit and hands-on skills;
- To provide personalized public services via Internet.

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Vision：Online and offline school environment



- Open, mobile, social, distributed and connected to the smart cognitive network and personalized development space.
- This ecological environment is not a fragmented learning space, rather a network connecting to the global community. Learning is not limited to classroom and school, but a lifelong, all-round and on-demand practice.

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Backgrounds & Aims

- Artificial intelligence is emerging from science fiction to everyday life, it continues to influence industries like consumer electronics, E-commerce, media, transportation, and healthcare.

Education —— Next Opportunity!

- Chinese government also announced its intentions to prioritize the development of AI as part of its national development plan.
- Provide an innovative platform for international research cooperation, understanding and investigating how AI could reinvent the future of education from both teaching and learning perspectives.

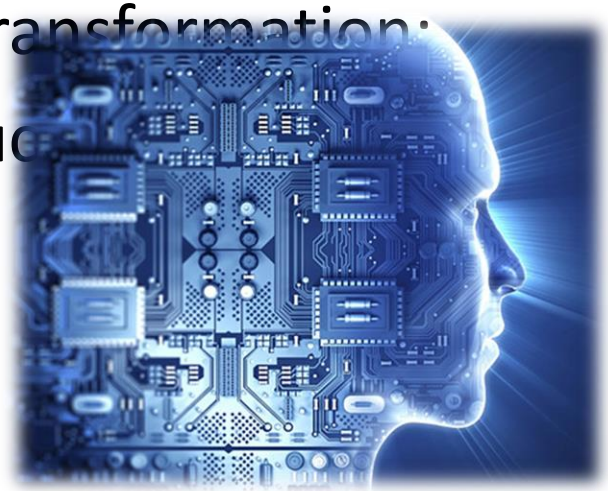
Scopes

- AI-driven knowledge base construction, knowledge graph construction and ontology construction;
- AI-driven knowledge tracing, educational data mining and learning analytics;
- AI-driven learner emotion recognition and affective computing;
- AI-driven new generation of student model and adaptive learning system;
- AI-driven automatic question generation, automatic question answering and automatic short answer grading;
- AI-driven problem-solving ability assessment;
- AI-driven student academic performance and achievement prediction;

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Scopes

- AI-driven recommender system for student career development;
- AI-driven intelligent teaching robot and agent;
- AI-driven interactive teaching with natural language processing techniques
- Ethics and law for AI-driven teaching and learning;
- Large scale educational data storage, processing and transformation;
- Any other relevant AI techniques applicable to the educ



Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Supports

- Up to US $ 50,000 grants, depending on the project
- AICFE will assign at least one researcher to collaborate with the grant recipients. We may also provide research engineers/assistant to conduct the system implementation.

Completion & Publication

- Reports
- Seminars
- Publish at least one journal paper (Indexed by SCI or SSCI)
- Publish at least one top conference paper
- Patents and system prototypes are also strongly encouraged

- Duration: 1-2 years

Proposal Submission Deadline

- July 30, 2017 (1st Stage)
- November 30, 2017 (2nd Stage)

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

More Information

- Handout in your bag
- Website: http://aic-fe.bnu.edu.cn/en/

- Contact: Sylvia Gao & Victor Lu
- Email: aitutor@bnu.edu.cn

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Objectives of SLP

SMART LEARNING PARTNER

- Data collection during the entire learning process
- Model construction for knowledge and capability
- Diagnosis and treatment for learning obstacles
- Identification and enhancement on disciplinary advantage

Sino-Finnish Joint Learning Innovation Institute 中芬联合学习创新研究院

Data Analysis Framework

Assessment
Assignment
Practice
Online learning

Clustering
Classification
Frequent mode
Outlier
Correlation analysis
Discriminant analysis
Comparison and summary
Trend analysis
Deviation analysis
Pattern discovery

Data collection

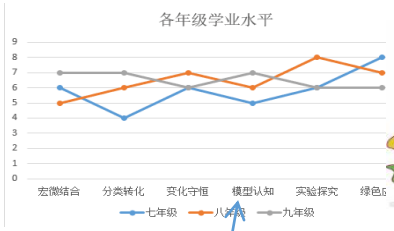
Datamation

Data mining

Modeling

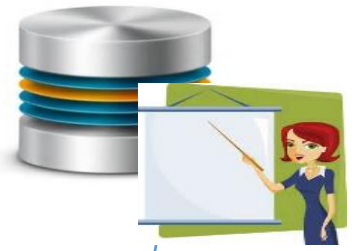
Online interaction
Works
Video recording
Classroom interaction
Mobile interaction
Wearable device
Intelligent device
Sensor network
Information system

Coding analysis
Text analysis
Discourse analysis
Pattern recognition
Voice recognition
Image analysis
Video analysis



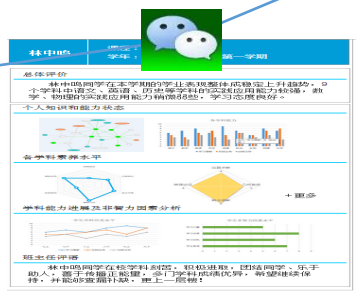
Education quality map

Service supplier



教育资源与服务统一战线

Intelligent recommendati engine



Individual diagnostic report

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

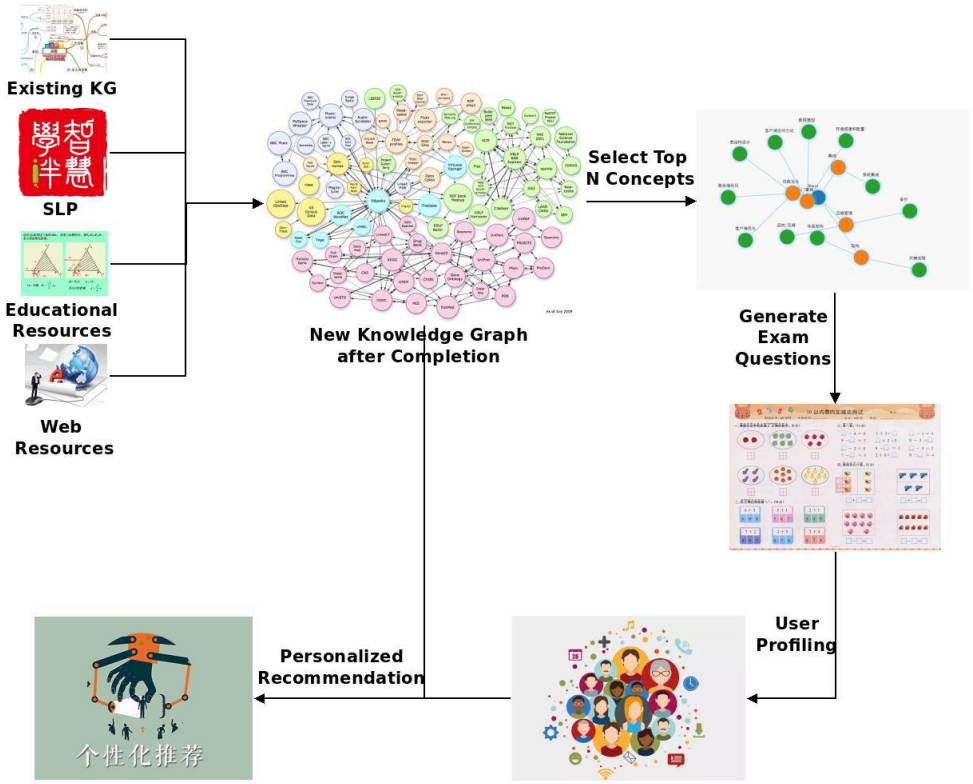
Research on Educational Knowledge Graph

Dr. Hepeng Cheng

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Educational Knowledge graph

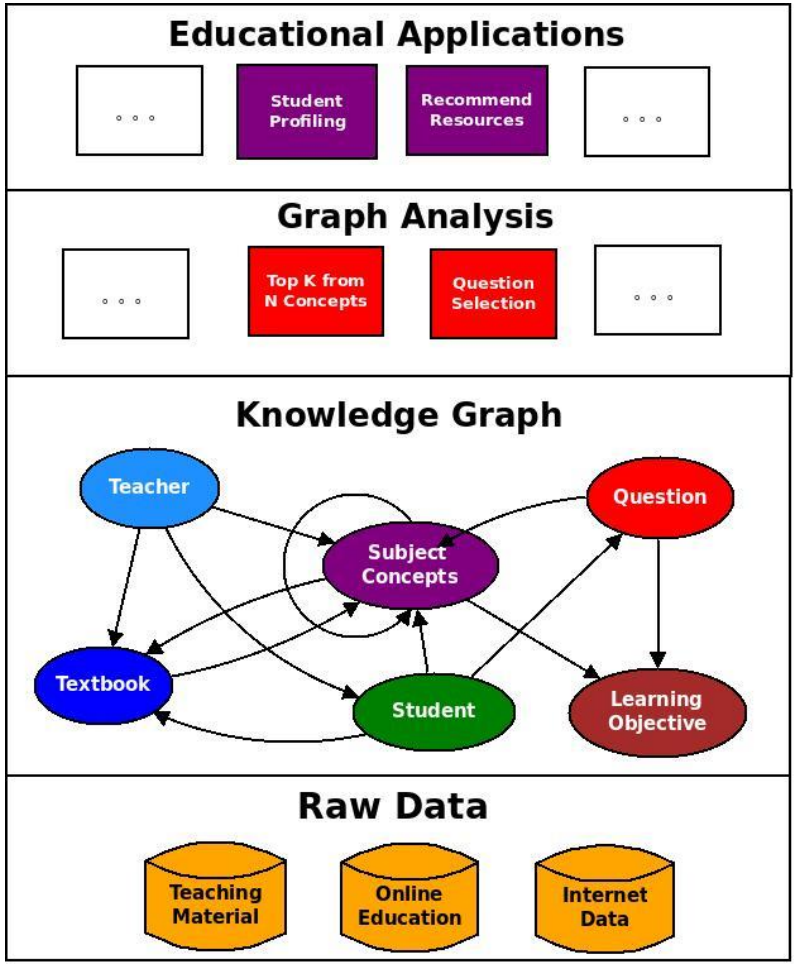
- **Objective**
  - To construct knowledge graph of K-12 education
- **Background**
  - Knowledge base of AI Tutor
- **Applications**
  - Knowledge state based student profiling
  - Intelligent personalized recommendation on learning resources
- **System Output**
  - Knowledge graph fused domain expertise and artificial intelligence
  - Automatic exam paper generation for given concepts
  - Student profile of knowledge states utilizing performance data
  - Personalized educational resource recommendation based on student profiles



Sino-Finnish Joint Learning Innovation Institute 中芬联合学习创新研究院

Educational Knowledge graph

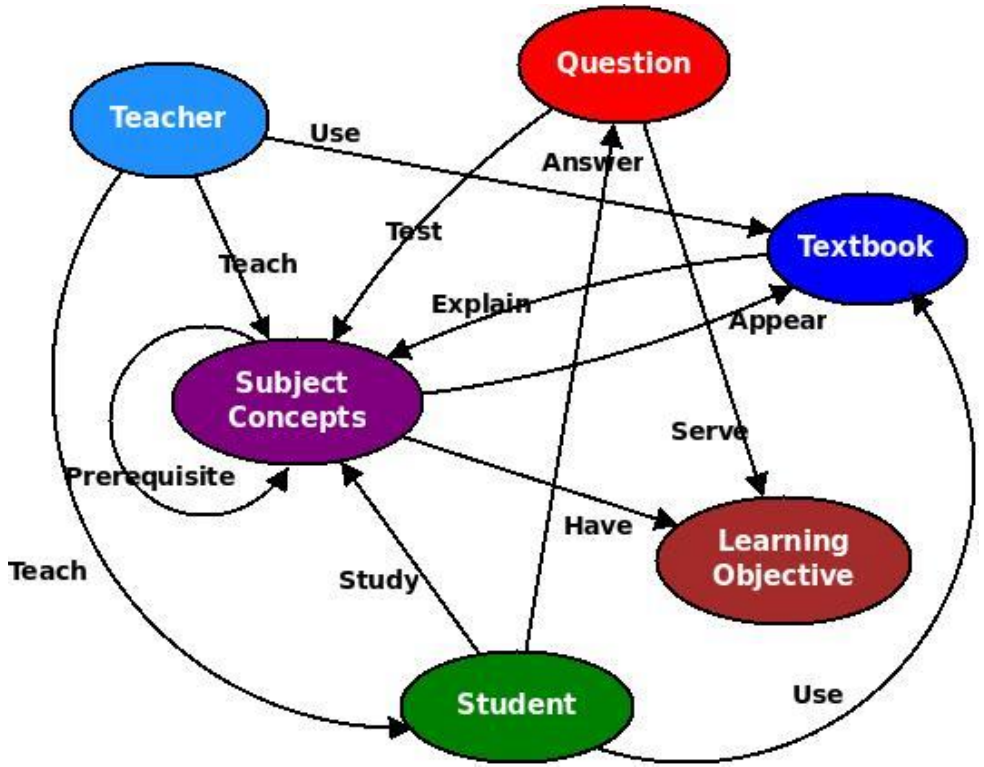
- **Objective**
  - To construct knowledge graph of K-12 education
- **Background**
  - Knowledge base of AI Tutor
- **Applications**
  - Knowledge state based student profiling
  - Intelligent personalized recommendation on learning resources
- **System Output**
  - Knowledge graph fused domain expertise and artificial intelligence
  - Automatic exam paper generation for given concepts
  - Student profile of knowledge states utilizing performance data
  - Personalized educational resource recommendation based on student profiles



Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Task 1: Knowledge Graph Construction

- Data Model



Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Task 1: Knowledge Graph Construction

- Objective
  - Fill in the content of knowledge graph according to designed data model, more specifically, include
    - Subject concepts and prerequisite relations between them
    - Linking subject concepts with textbooks and questions
    - Linking subject concepts with learning objectives
    - Linking to students and teachers
- Data Sources
  - Traditional teaching material: textbooks, lecture notes, curriculum standards
  - Online education platform: learning log, teacher-student interaction, forum data
  - Internet data: Wikipedia data
- Output
  - Knowledge graph that fused domain expertise and artificial intelligence

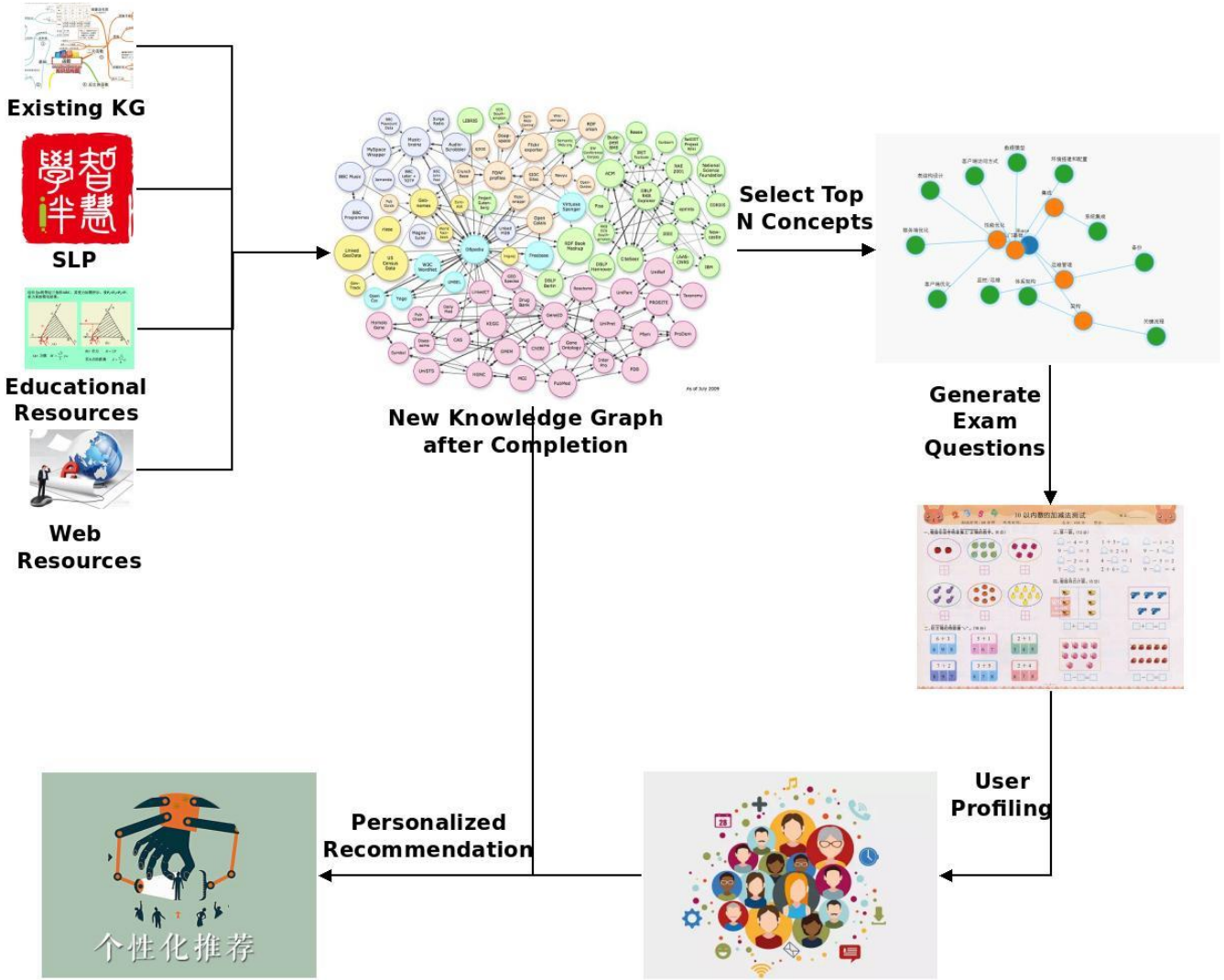
Task 2: Knowledge Graph Analysis

- Objective
  - Use a small set of questions to examine students' knowledge states of a large set of subject concepts
- Subtasks
  - Subset selection
    - Find out subset of subject concepts to covert the entire set of given subject concepts
  - Paper generation
    - How to build a paper with given subject concepts and their related questions?
- Output
  - An algorithm to generate exam paper based on given subject concepts for testing

Task 3: Educational Application

- Application 1: Student Profiling
  - Objective
    - Monitor/Represent students' knowledge states based on knowledge graph, students' particulars, performance data
  - Challenge
    - Performance data is not continuous due to limited number of performance data, in which case we need to predict students' performances on those subject concepts without performance data
  - Output
    - Students' profiles of knowledge states
- Application 2: Smart Recommendation
  - Objective
    - Learning resources recommendation based on student profiles as well as learning objectives
  - Challenge
    - Matching and coverage between subject concepts and questions
    - Matching and coverage between subject concepts and learning resources
  - Output
    - Recommended resources

System Workflow



Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Accomplishment and Collaboration

- Accomplishment
  - Half done with task 1
  - Finished:
    - Subject concept extraction (first round, may iteratively update in the future)
    - Prerequisite relations identified manually
    - Linked learning objectives with certain key subject concepts
    - Linked subject concepts with several sets of exam questions
  - Remaining:
    - Linking subject concepts with textbooks
    - Linking subject concepts with more questions
    - Linking subject concepts with teachers and students
- Potential Collaboration
  - Knowledge graph construction: share data and resources to enrich our knowledge graph content
  - Knowledge graph analysis: work on certain graph analysis together
  - Educational application: develop certain educational applications on top of knowledge graph and analysis

Automatically question generation based on semantic network

Dr.Lishan Zhang

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

Common question generation techniques

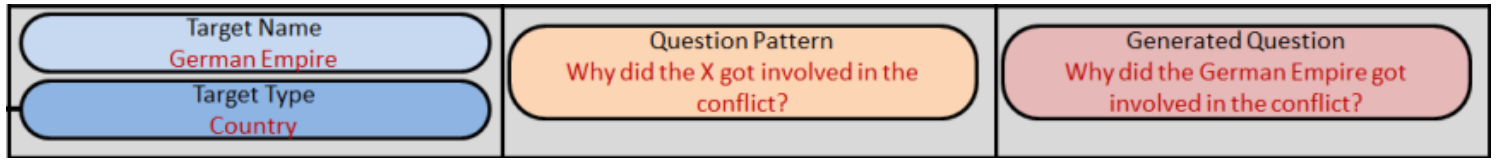
- Generation based on plain text
  - For example

John bought some fruits

```
(ROOT
(S
        (NP (NNP John))
                (VP (VBD bought)
                        (NP (DT some) (NNS fruits)))))
```

-> Who bought some fruits?
-> What did John buy?

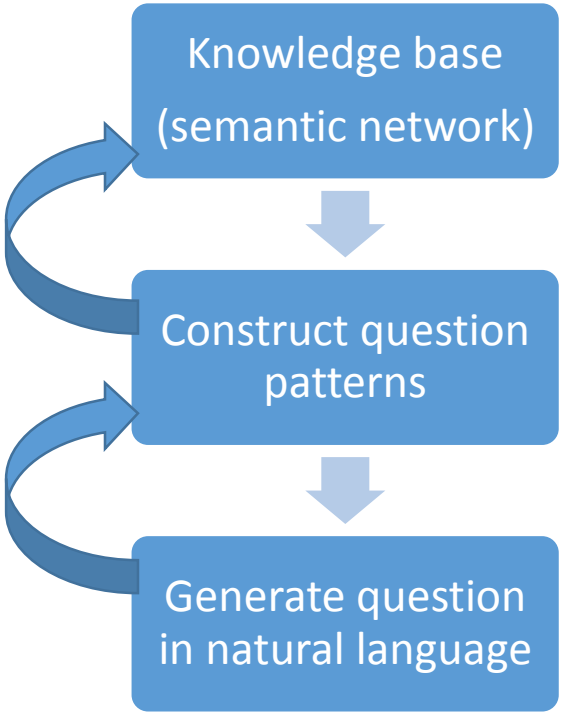
- Generation based on a semantic network or ontology for a specific domain (We adopt this methodology)
  - For example.



The Workflow for question generation

Evaluate and improve question patterns &
Improve the standards of knowledge base

Evaluate and improve the question patterns
by looking at the generated questions



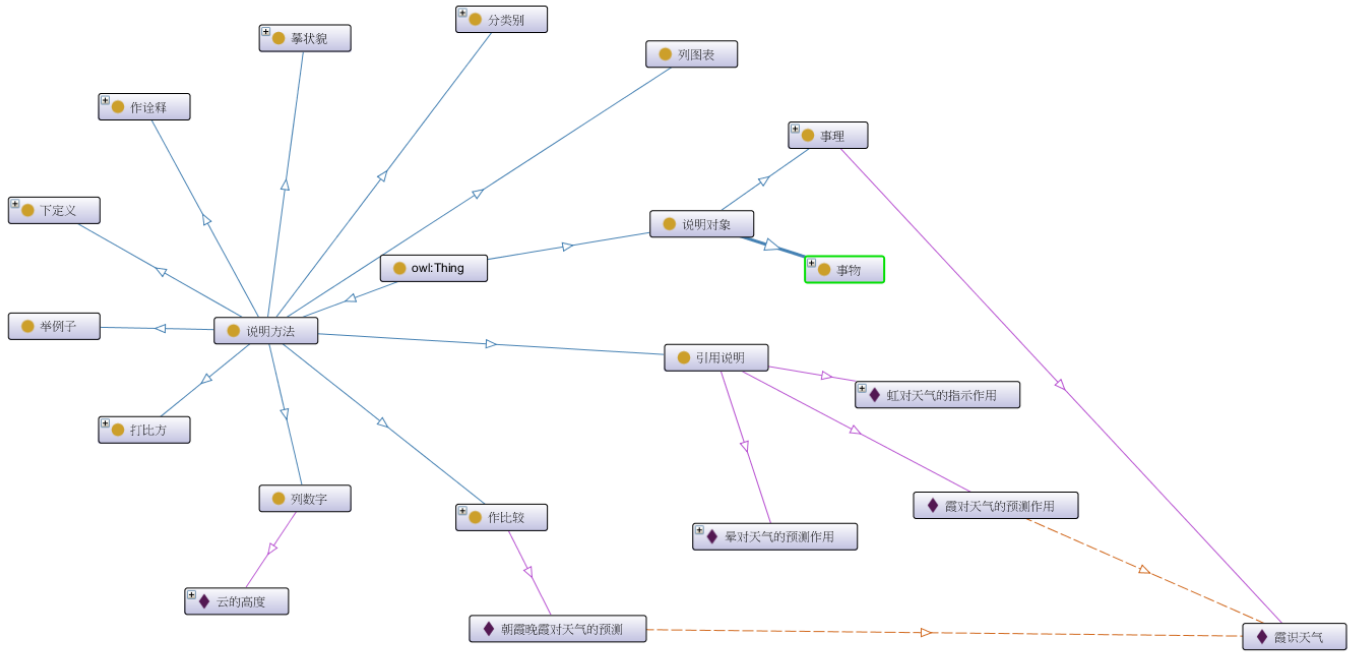
Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

The domains

- Chinese reading comprehension
  - Expository text reading in specific
  - Aim for improving students' understanding on the text
- Photosynthesis in Biology
  - Aims for helping teachers generate shallow questions
  - Aims for assessing students' understanding on basic concepts

Question generation for expository text reading comprehension

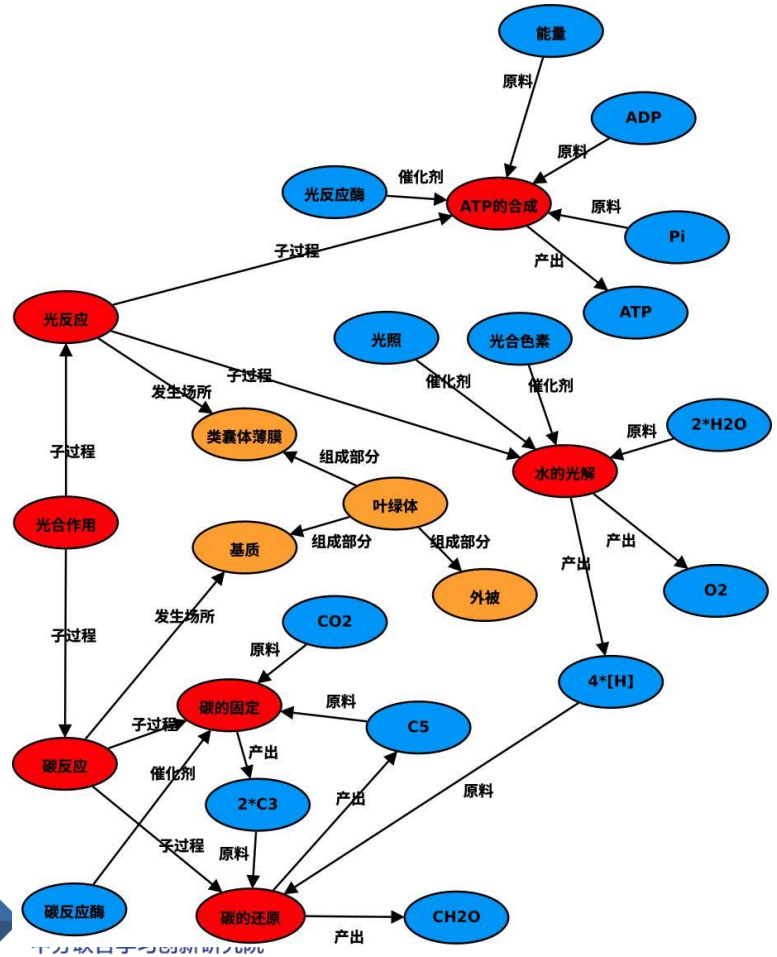
- Text code schema:
  - The object of the expository text is classified into two types
  - The way to describe the object is classified into ten types
- The plain text is transformed into a semantic network:



Sino-Finnish Joint Learning Innovation Institute
中芬联合学习创新研究院

Question generation in photosynthesis

- Each concept is classified as process or instance
- The knowledge in this domain is transformed into the semantic network:



To generate questions like

What does photosynthesis produce?

How is light used in photosynthesis?

Where does photosynthesis take place?

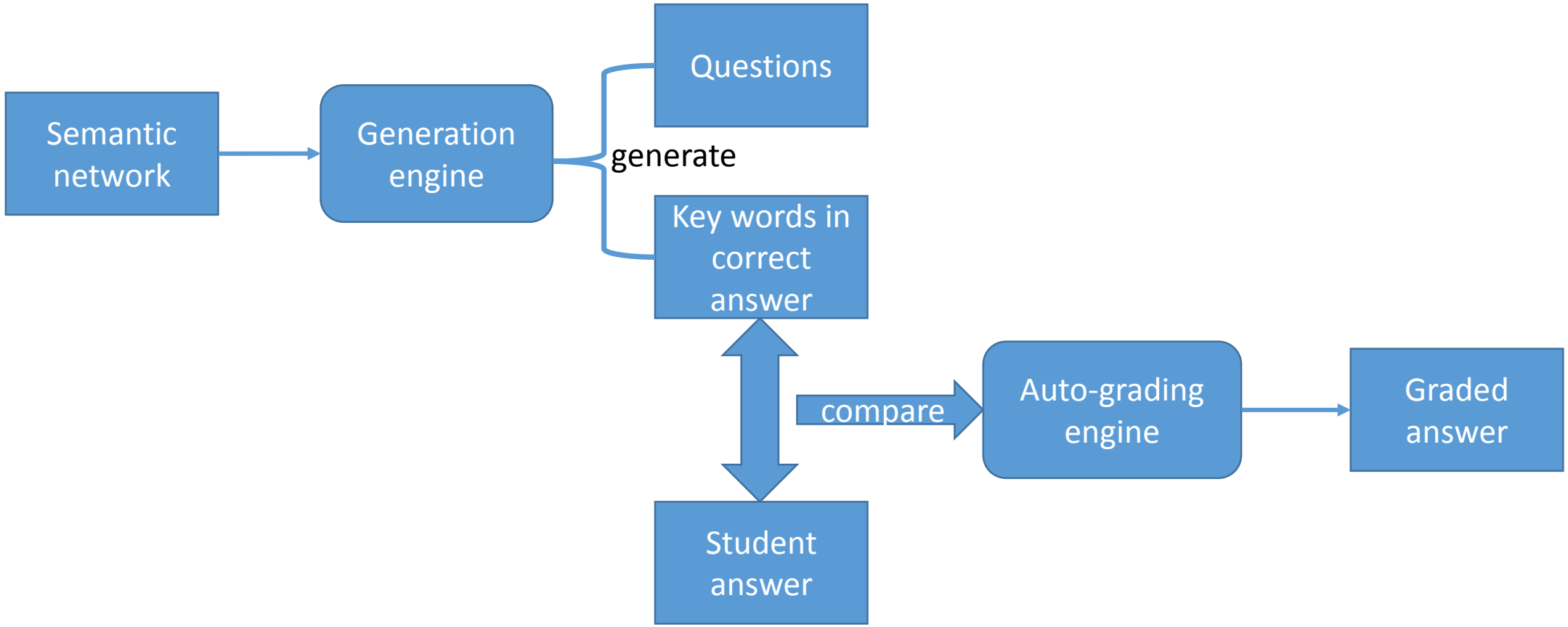
...

Technologies being used

- OWL standard is adopted to describe the semantic network
- Jena API as well as SPASQL is used for accessing the coded semantic network
- The generation program is being implemented with Java
- The program accesses the semantic network recursively to find out all the relations fitting for the question pattern.

Connection with auto-grading component

- Both question and its answer can be generated from the semantic network. So it can facilitate student answer grading.



Expected results

- By having teacher authorized a semantic network, our system can automatically generate questions to auto-grade students' answers, assess students' competence, feedback to students, and adaptively select the next question.

Automated Assessment System for Short Answer Questions

Dr.Xi Yang

Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

We Need Autamatic Grading

1. Assessing students' acquired knowledge is one of the key aspects of teachers' job. Assessments are important for teachers as these provide them insights on how effective their teaching has been. However, assessment is a monotonous, repetitive and time consuming job and often seen as an overhead and non-rewarding.

2. Consequently, use of open-ended questions that seek students' constructed responses is more commonly found in educational institutions. They reveal students' ability to integrate, synthesize, design, and communicate their ideas in natural language.

3. With the increase of e-learning, MOOCs, online testing automatic grading has aroused more critical discussions.

Break Through on Reading Comprehension

- Different question type is for different level of cognitive skill
- Different question type is corresponding to different openness level
- There are few researches in automatic reading comprehension grading

Datasets

Collecting data is a significant part for our researches.

•**Chinese Data** We organize two experienced teachers to label the Chinese answers individually and made agreements finally.

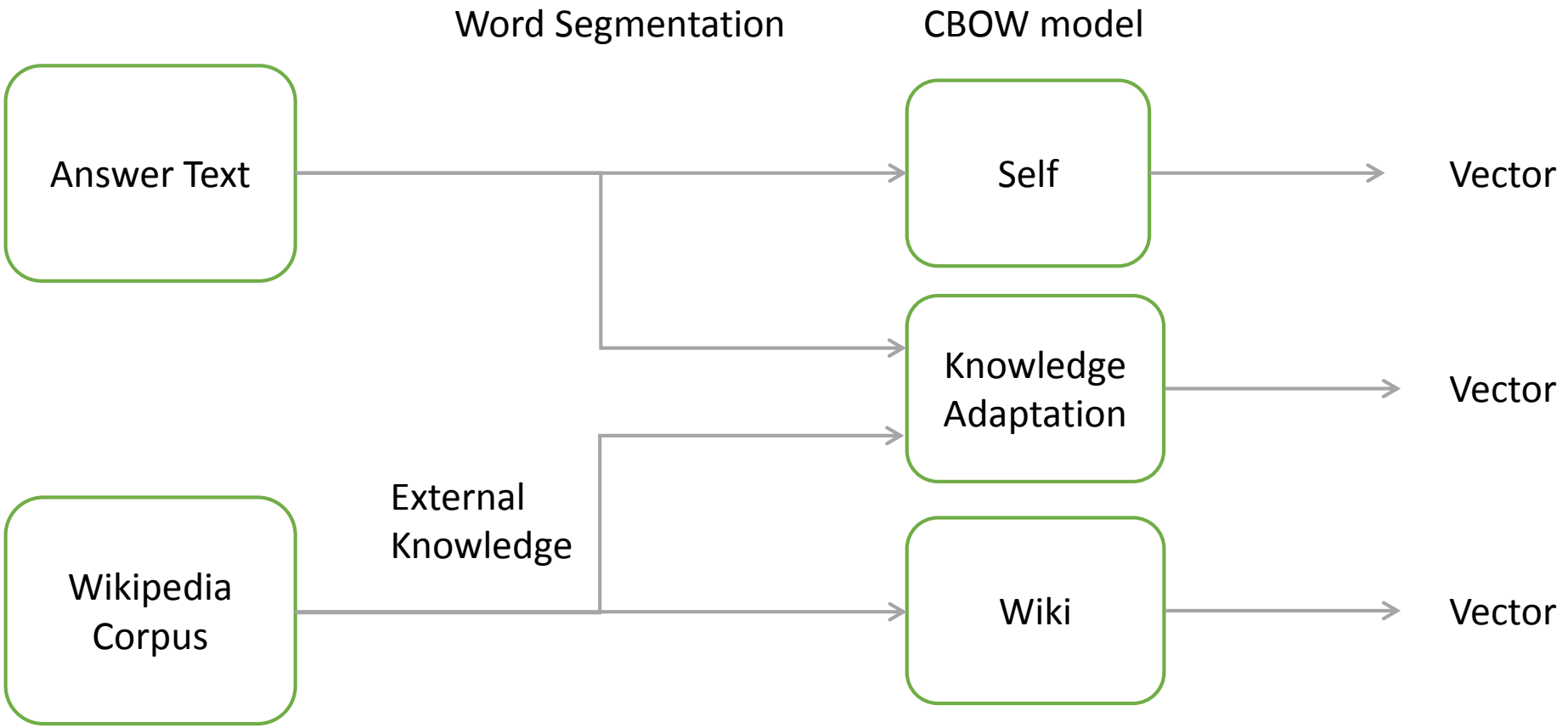
•**English Data** We selected five datasets in Kaggle Automatic Student Assessment Prize: Short Answer Scoring(ASAP-SAS) based on reading comprehension definition.

Data Overview

Problem	Avgword	Total Samples	Score Level	Language
CRCC1	39	2579	0-4	Chinese
CRCC2	33	2571	0-2	Chinese
CRCC3	26	2382	0-3	Chinese
CRCC4	27	2458	0-4	Chinese
CRCC5	31	2538	0-3	Chinese
ASAP-SAS3	47	2297	0-2	English
ASAP-SAS4	40	2033	0-2	English
ASAP-SAS7	41	2398	0-2	English
ASAP-SAS8	52	2398	0-2	English
ASAP-SAS9	49	2397	0-2	English

Algorithm

Answer Preprocessing & Word Embedding



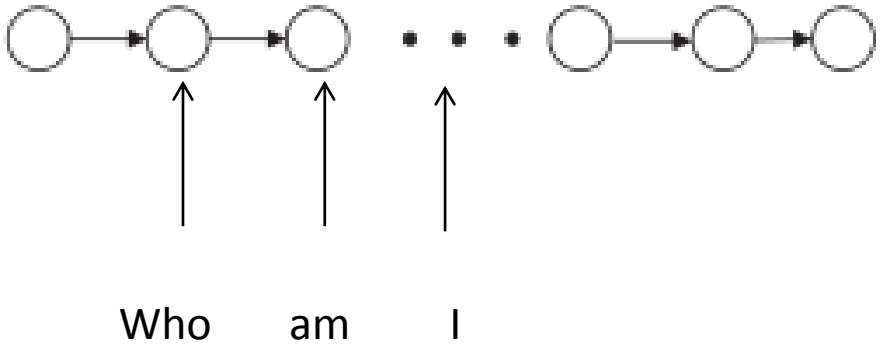
Sino-Finnish
Joint Learning Innovation Institute
中芬联合学习创新研究院

LSTM Extract Semantic Information

Standard Bag-of-words Model

ignore the word order

LSTM based Deep Sequence Model



consider the word order and disposal word sequence

Experiment and Conclusion

- Ten datasets (5 Chinese & 5 English)
- Two baselines
  1. Logistic Regression
  2. Support Vector Machine
- Evaluation: Accuracy

Results on Accuracy

	DJDT	HSVM	LSTM+self	LSTM+web corpus	LSTM+KA
CRCC1	0.5106	0.5482	0.5979	0.5805	**0.6134**
CRCC2	0.6036	0.6585	**0.7487**	0.7242	0.7379
CRCC3	0.8564	**0.8862**	0.6511	0.8061	0.8229
CRCC4	0.5020	0.5867	0.5533	0.5725	**0.5911**
CRCC5	0.7574	0.7660	0.6942	0.7443	**0.7738**
ASAP-SAS3	0.4789	0.4698	0.4806	0.4885	**0.4898**
ASAP-SAS4	0.6385	0.7358	0.4550	0.7688	**0.7742**
ASAP-SAS7	0.6343	0.6626	0.5605	**0.6684**	0.6493
ASAP-SAS8	0.5234	0.5988	0.5868	0.6222	**0.6322**
ASAP-SAS9	0.6237	0.6442	0.6834	0.6458	**0.6926**
Ava accuracy	0.6126	0.6557	0.6011	0.6623	**0.6748**

Analysis

- HSVM is a relative better automatic reading comprehension grading model in baselines.
- The statistic machine learning models still work in some reading comprehension grading tasks without rubrics. The pretrained word vectors is limited for LSTM approaches. And the vectors training performance is influenced by the volume of datasets so the word embedding may not perform better when only use the student answers.
- More importantly, the experiment results also proved that transfer external knowledge for word embedding through knowledge adaptation can help impove the performance of model.

Conclusion

- We propose a deep learning based method for automatic Chinese reading comprehension grading. Our method does not rely on any target answer due to the fact that target answer is not always available for most open-ended reading comprehension questions.
- In our framework, CBOW and LSTM are combined and extract semantic information automaticly and effectively consider the word orders in student response.
- Additional, through knowledge adaptation, the external knowledge is transferred to present corpus by utilizing fine-tuning technique. Experiments on ten datasets, demonstrate the performance improvement by introducing of external knowledge.

THANK YOU!