

Forst: Question Answering System for Term and Essay Questions at NTCIR-13 QA Lab-3 Task

K. Sakamoto (YNU, NII), M. Ishioroshi (NII), Y. Fukuhara (YNU), A. Iizuka (YNU), H. Shibuki (YNU), T. Mori (YNU), N. Kando (NII)

Motivation

- We tackled the **term question task** and the **essay question task** including the **evaluation-method subtask**
- Our systems for the term question task and the essay question end-to-end subtask are **successors of our systems at the QA Lab-2**

Knowledge Source

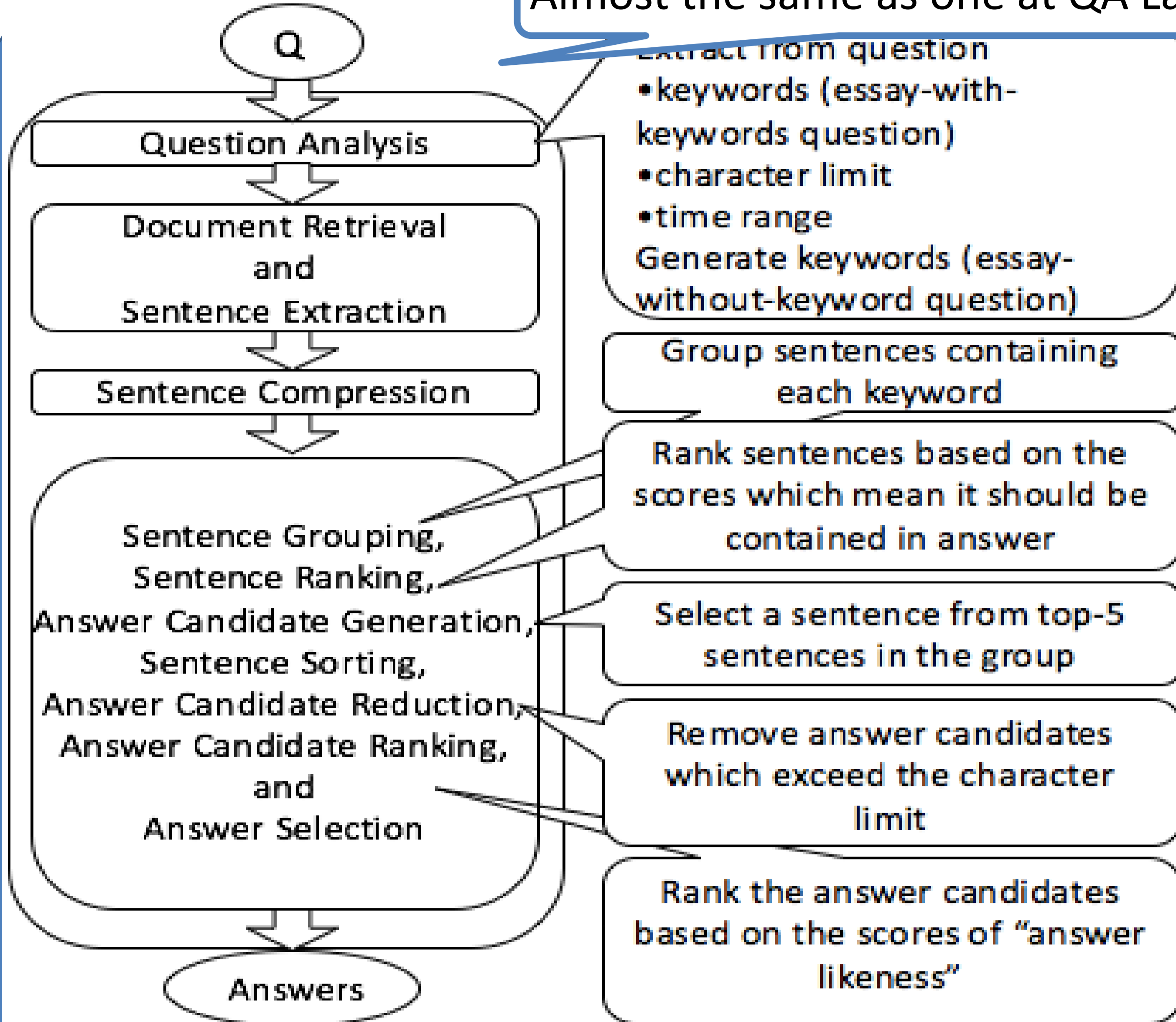
- 4 textbooks (Given in the task)
- Q & A collection (4,324 pairs)
- Glossary (6,081 words)
- Japanese thesaurus (about 300,000 entry words)
- world history event ontology (Given in the task)

Instance data from ontology

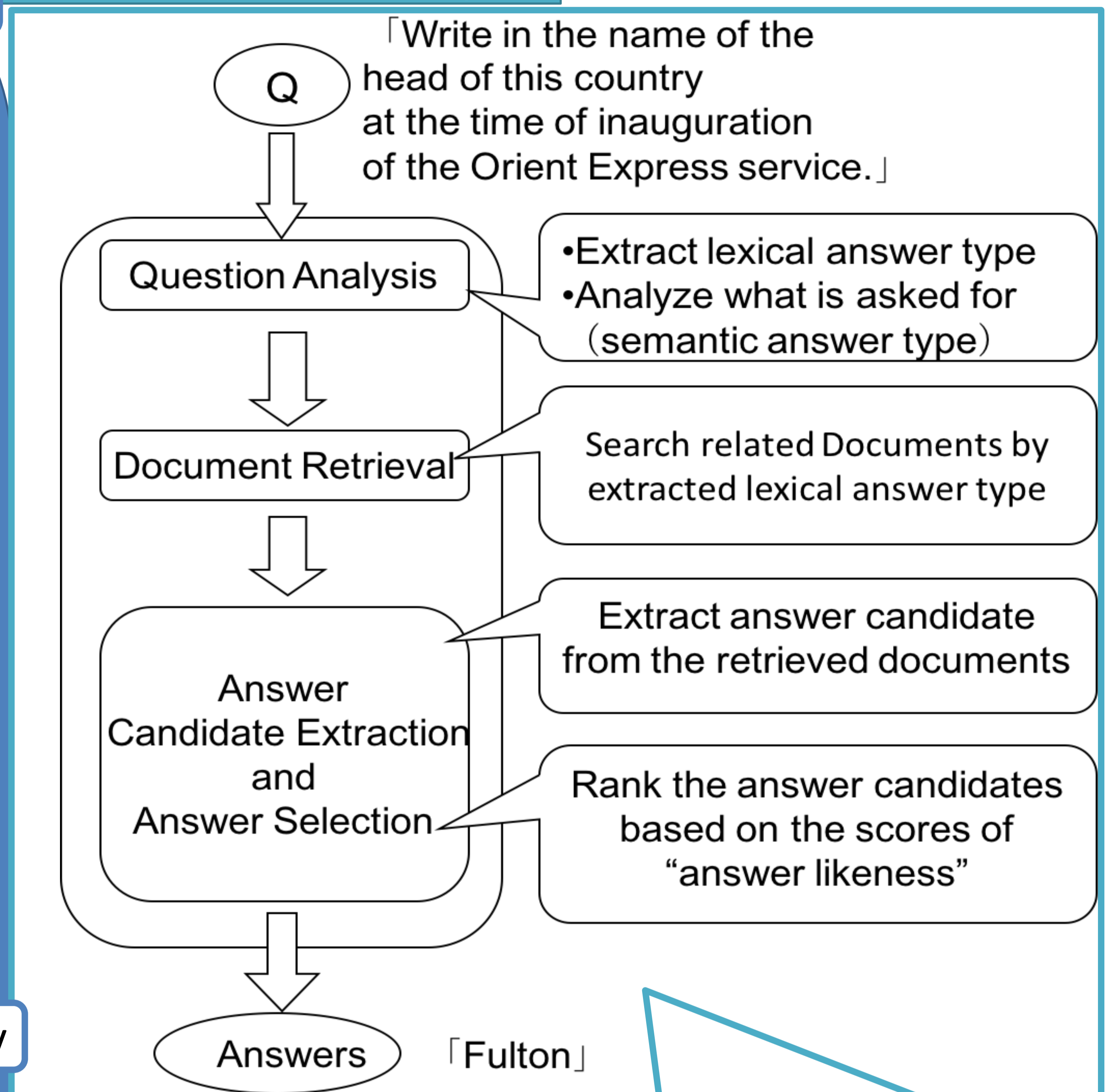
国名表現クラス、サブクラス、国名表現、開始年、終了年、異表記、その他の情報 (複数ある場合は@で区切る)
 技術、文字、アラビア文字、アラビア文字、インド
 技術、文字、キリル文字、900 頃、
 技術、語、グレゴリウス語、1582、グレゴリオ語、ローマ@グレゴリウス 13 世人物、発明家、アークライト、1732、1782、グレートブリテン王国

Essay type answering

Almost the same as one at QA Lab-2

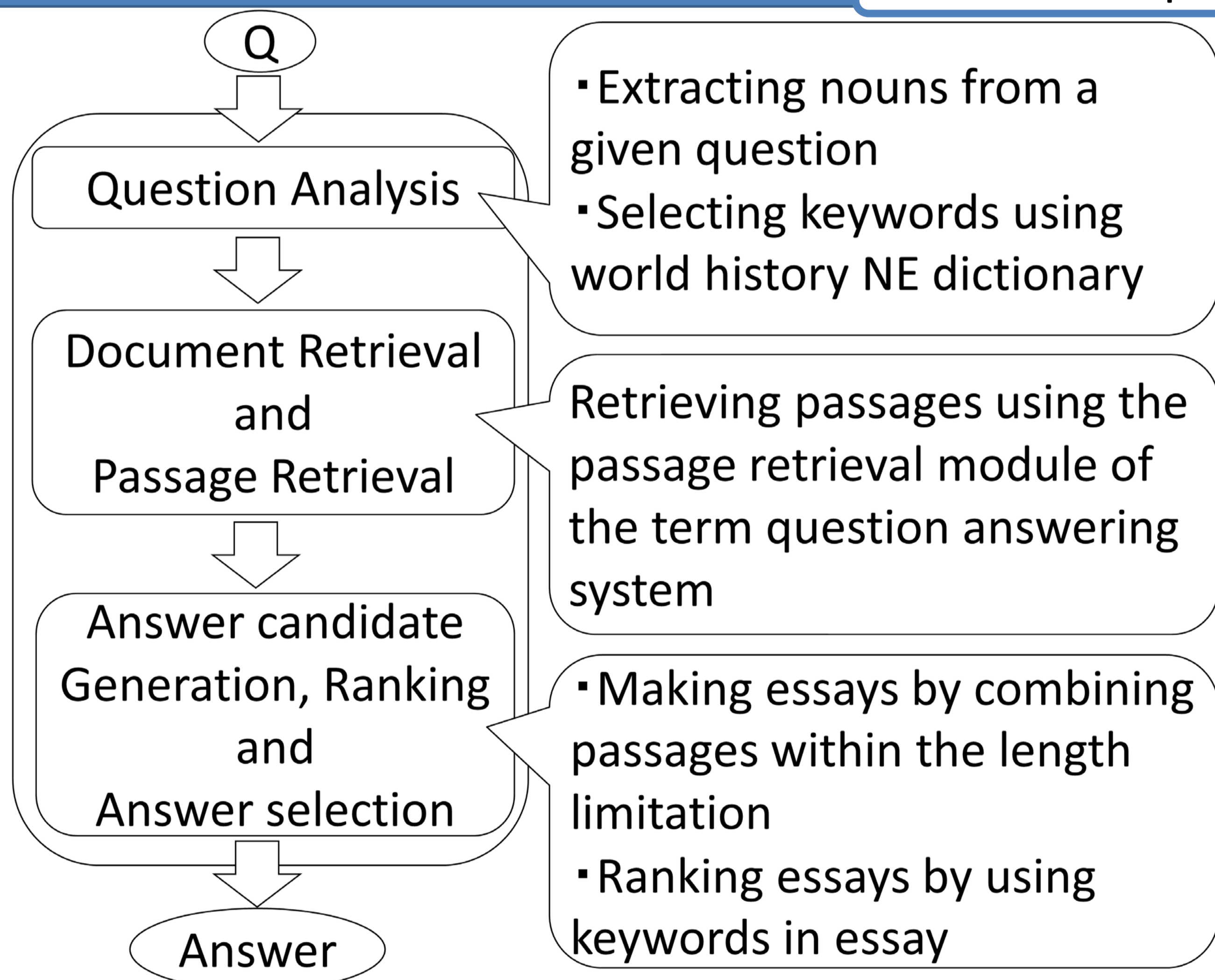


term type answering



Essay type answering (the second system)

Focus on simple essay



Changes since the QA Lab-2

- Using **keyword importance**
 - The later keywords appear in a question, the more emphasized.
- Extending dictionary** for NE of world history
- Adding decision rules** for question types
- Using **majority decision score** for answer selection

The concept is use of **'implicit keywords'** that are question focuses but not stated positively.

Evaluation method

- Based on **world history terms**
- Based on a given set of **gold standard nuggets**
 - The system count the **number of terms (nuggets)** in a essay
 - The count is regarded as the **essay score**

Results

Evaluation method

Term	Accuracy	priority	Spearman's Rho	Kendall's Tau b
Phase1	0.397	1	0.427	0.334
Phase2	0.273	2	0.596	0.534
Phase2		1	-0.071	-0.049
Phase2		2	0.404	0.360

Essay	priority	Human (Complex essay only)	ROUGE-1
Phase1	1	0.011	0.0523
Phase1	2		0.0698
Phase2	1	0.0339	0.0385
Phase2	2		0.0680

The second system only answered **simple essay**

Conclusion

- We participated in all phases of the term question task and the essay question task in Japanese
- Although the **changes since the QA Lab-2** did **not** bring the major **improvement**
- Using **'implicit keywords'** extracted from question texts makes the **results better**
- The evaluation method using **gold standard nuggets** **achieved the best results**