

The Interface Linguistics / Knowledge Representation

Question Answering II

Hinrich Schütze, IMS

University of Stuttgart

Wintersemester 2004/05, Jan 24, 2005

Today's topic

- Question answering II
- Redundancy vs. Linguistic analysis

Sources

Li and Croft: The impact of syntactic evidence on the effectiveness of question answering

Redundancy vs. Linguistic analysis

- Pros and Cons?

Baseline system: Heuristic score

$$heu_score = N + 0.5 * Sm + N/W + 0.5/D \quad (1)$$

where four heuristics are considered: the number of matching query words (N), whether the matching words are in the same sentence ($Sm=0/1$), the size of the best matching window (W), and the distance between an answer candidate and the center of the best matching window (D).

Table 1. The Scoring Algorithm in the Baseline QA System

1. Do the following for each answer candidate in the top 10 passages;
2. Initialize SCORE to 0;
3. Match each query word with words in each passage. Let N stand for the *number of matching words*, then
 $SCORE = SCORE + N$;
4. Check whether all matching words in the passage are in a *single sentence*. If yes, then
 $SCORE = SCORE + 0.5$;
5. Locate the *best matching window* in the passage and calculate the size of it, and the score is updated as
 $SCORE = SCORE + N / \text{size of the best matching window}$;
6. Locate the answer candidate in the passage and calculate the *distance* between the candidate and the center of the matching window in token offset. The final heuristic score is updated as
 $SCORE = SCORE + 0.5 / \text{DISTANCE}$

Question parse tree

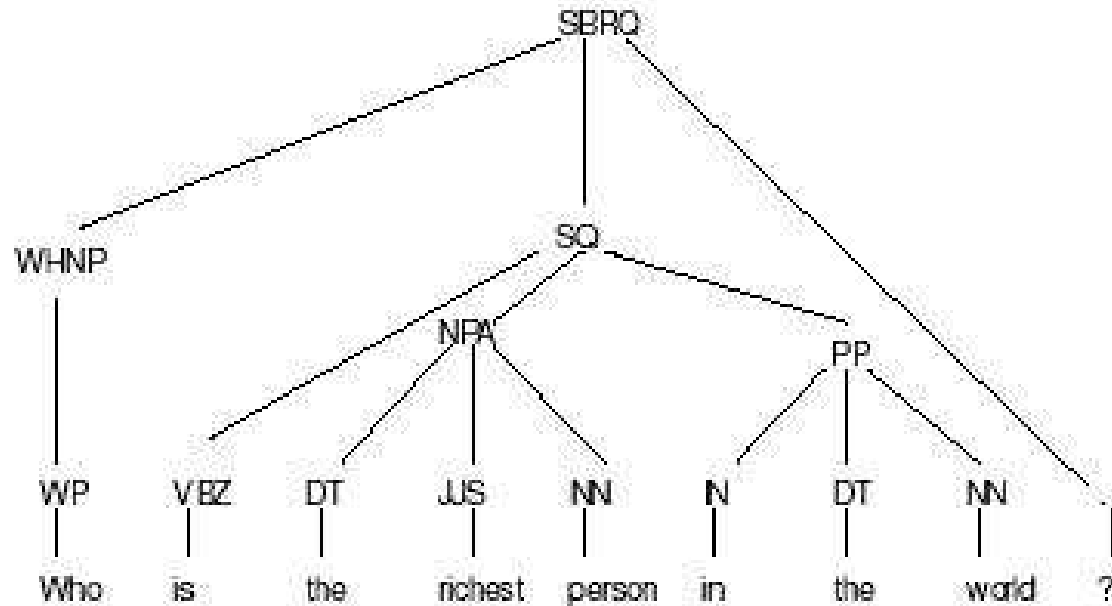


Figure 1. Parsing tree of the question “Who is the richest person in the world?” The actual output from the BBN parser we used is a string that can be easily rebuilt into the tree structure of the question.

Syntax can help: Example

631, “Who won the Nobel Prize in literature in 1988”, the best passage that has the correct answer is as follows:

“After Naguib Mahfouz, who won the 1988 Nobel Prize in literature, Abdel-Kuddous was among the best-known novelists in the Arabic language.”

Phrases are important

Table 2. A question and top three passages in the documents returned by INQUERY

| | |
|-----------|--|
| Question | <i>Who is the richest person in the world?</i> |
| Passage 1 | <i>Although tops in the U.S., Mr. Walton is the sixth-richest person in the world.</i> |
| Passage 2 | <i>Once the richest black person in the world, Baker was destitute shortly before her death. She died in her sleep on the second night of a phenomenally successful comeback show in Paris.</i> |
| Passage 3 | <i>As well as being the richest person in the world, Sir Hassanal lives with his relatives in the world's biggest palace _ a complex of buildings built with 38 types of marble on a 300-acre hill near the Brunei River. In case friends decide to stay over, it has 1,778 rooms and 257 toilets.</i> |

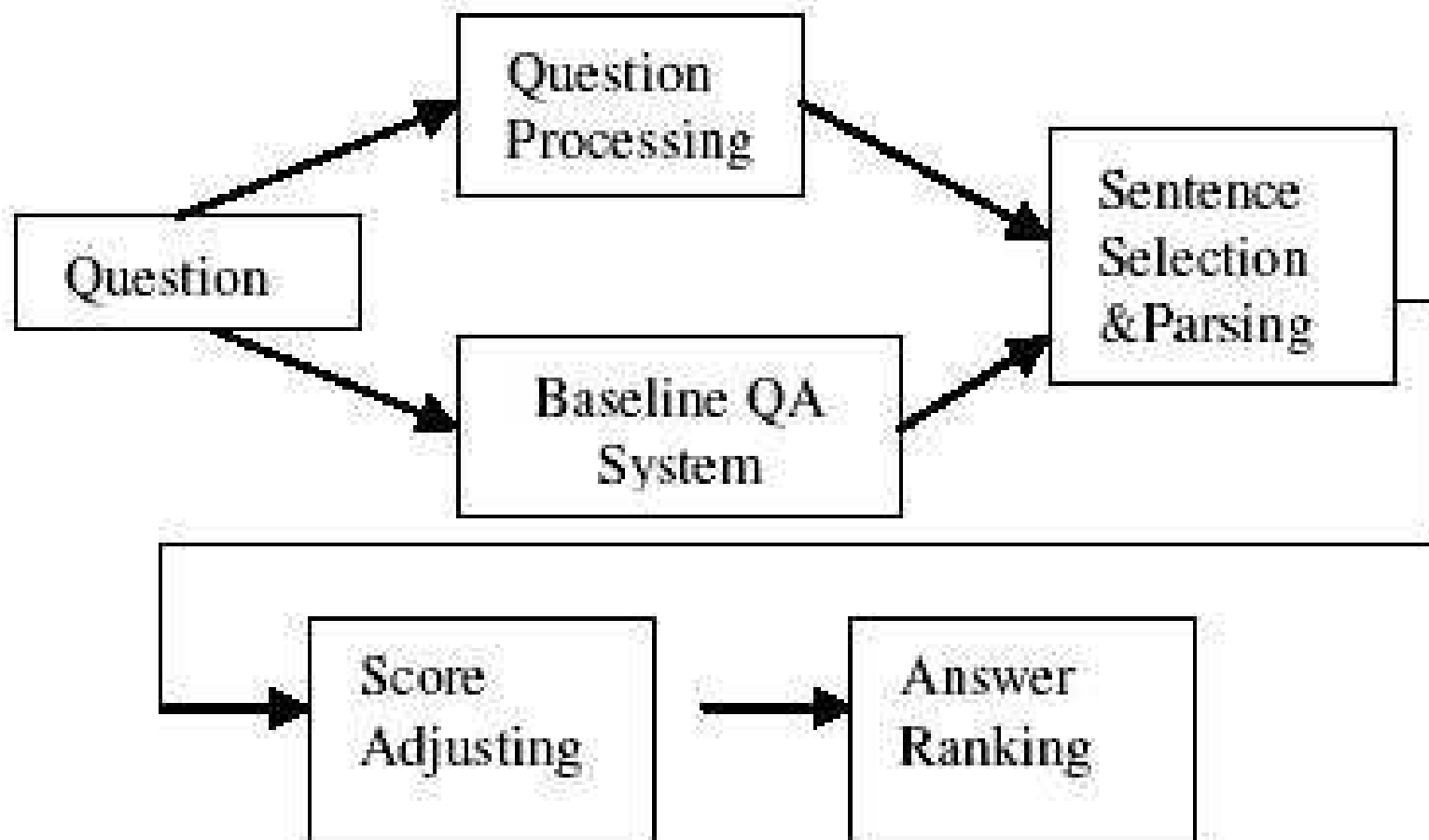


Figure 2. Framework of the New QA System

Step 2: Passage selection

$$heu_score^* = N + 0.5 * Sm + N/W + 0.5/D + 0.5 * Sc \quad (2)$$

where $Sc = 1$ if the candidate and all the matching words are in the same sentence, otherwise 0.

Syntactic clues

Table 3. Six syntactic factors in the new QA system

| | |
|----------|--|
| Factor 1 | Match the sentence with the phrases extracted from the question. If a longer phrase is matched, then the short phrases within it will not be further considered. F1 = the size of total matched phrases/the size of the question. |
| Factor 2 | Consider the distance between the answer candidate and the main verb in token offset. F2 = the distance between the answer candidate and the main verb. |
| Factor 3 | For "PERSON", check the relationship between the answer candidate and the main verb in the sentence to see if it is consistent with the relationship in the question. F3 = 1 if factor 3 is satisfied, 0 otherwise. |
| Factor 4 | For "LOCATION" questions, check the possessive formats such as, "Venezuela's Orinoco" and "Orinoco in Venezuela". F4 = 1 if factor 4 is satisfied 0 otherwise. |
| Factor 5 | For "LOCATION" and "DATE" questions, check whether the candidate is inside a prepositional phrase and modifies the main verb. F5 = 1 if factor 5 is satisfied 0 otherwise |
| Factor 6 | For "PERSON" questions, check whether the candidate and all query words are inside a NPA (adjective noun phrase). F4 = 6 if factor 6 is satisfied 0 otherwise. |

Syntactic and final score

$$\textit{syn_score} = 1.0 * F1 + 0.5 * F2 + 0.5 * F3 + 1.0 * F4 + 1.0 * F5 + 1.0 * F6$$

$$\textit{Final_score} = \textit{heu_score}^* + \textit{syn_score}$$

Evaluation

Table 4. Experimental Results

| Question Type | All | Person | Location | Number ¹ | Date | Organization |
|------------------------|-------|--------|----------|---------------------|--------|--------------|
| $N_{\text{questions}}$ | 162 | 57 | 56 | 15 | 25 | 9 |
| M-MRR ² | 0.690 | 0.686 | 0.668 | 0.650 | 0.778 | 0.667 |
| Sift-MRR ³ | 0.744 | 0.775 | 0.753 | 0.724 | 0.690 | 0.667 |
| Change | 0.054 | 0.089 | 0.085 | 0.074 | -0.088 | 0 |
| Change of % | 7.8% | 13.0% | 12.7% | 11.4% | -12.7% | 0 |
| N_{improved} | 32 | 12 | 14 | 4 | 2 | 0 |
| $N_{\text{decreased}}$ | 14 | 3 | 4 | 1 | 6 | 0 |

¹**Number**: includes “NUMBER”, “MASS”, “MONEY”, “LENGTH”, and “PPERIOD” question types. ²**M-MRR** stands for the mean reciprocal rank using Marsha heuristics alone. ³**Sift-MRR** stands for the mean reciprocal rank incorporating syntactic information.

Evaluation: Impact of syntax

- Total number of questions: 162
- Unchanged/changed: 116/46
- Better performance: 32/46
- Worse performance: 14/46
- Decrease for DATE due to system problems?

Improvement due to syntax

Question 249: “Where is the Valley of the Kings?”

The sentence having the correct answer: “The newspaper said the remains have not been disturbed since they were sent to the gardens in 1932 by Howard

Carter, who discovered the Valley of the Kings at Luxor, Egypt in 1922.”

Example of failure

Question 851. “When did Mount St. Helens last erupt?”

Passage 1 and passage 2 are two passage candidates.

Passage 1: “Mount St. Helens could erupt again at any time,” said Don Swanson, scientist in charge at the USGS observatory in Vancouver Wash. Throughout its recorded history, Mount St. Helens has had active periods that lasted for years with relatively short spans of inactivity. Before the 1980s, the last eruptive period was from 1800 to 1857, with intermittent periods of quiet lasting months or years, according to the USGS western region office in The volcano's most recent eruptions have been quiet, dome-building affairs, in which the mountain pumps out thick lava to increase the size of the crater dome.”

Passage 2: “Mount St. Helens, historically one of the Cascade Range's most active volcanoes, had not erupted since 1857.”

Logistic regression results

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n \quad (7)$$

$$p = \text{Exp}(y) / (1 + \text{Exp}(y)) \quad (8)$$

where p is the probability that an answer candidate is a correct answer. x_1, x_2, \dots, x_n are independent variables.

Table 5. Performance using logistic regression (LR)

| Performance | All | Person | Location | Number | Date | Organization |
|-------------|-------|--------|----------|--------|-------|--------------|
| M-MRR | 0.690 | 0.686 | 0.668 | 0.650 | 0.778 | 0.667 |
| Sift-MRR | 0.744 | 0.775 | 0.753 | 0.724 | 0.690 | 0.667 |
| LR | 0.717 | 0.678 | 0.763 | 0.606 | 0.753 | 0.759 |