Maksel Language Brackalling

Kallie Bracken

&

Maks Cegielski-Johnson
Maksel Language Brackalling

Kallie Bracken

&

Maks Cegielski-Johnson
How Our System Works

What Did & Didn’t Work Well
How Our System Works

What Did & Didn’t Work Well
Input Files

Processing

Get Sentence Candidates

Weight candidates by
- Word Overlap
- NP Overlap
- VP Overlap
- Other: Named Entities, etc.

Question Classification

Choose Best Candidate

Best Sentence

Increase Precision

Final Answer

Parsing & Preprocessing Input

Question & Story Sentences
ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate using:
- Pure overlap
- NP Overlap
- VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.

For Who and Where questions increase PRECISION; Return only “person,” “organization,” and “location” NEs from best answer.

WHEN QUESTIONS
All sentences with “date” and “time” NEs are automatically weighted as candidates.

PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS
If (question contains: “How much/many/long/.../often/tall/old…”):
→ Sentences with numbers (digits or words) are weighted as candidates.
→ PRECISION: Only number entity from best candidate sentence is returned.

Else: Use Enhanced Overlap Approach

Expansion/Question as Set
- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Final Answer

Increase Precision
Remove question words from answer

Best Answer
ENHANCED OVERLAP APPROACH

\[ \text{enhancedOverlap}(\text{Question question}, \text{sentences}, \text{NPWeight}, \text{VPWeight}) \]

**FIND CANDIDATES** – Weight each sentence as potential candidate using:
- Pure overlap
- NP Overlap
- VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.

**ASSIGN BEST ANSWER** to be sentence candidate with highest weight.
For Who and Where questions increase PRECISION; Return only “person,” “organization,” and “location” NEs from best answer.

**WHEN QUESTIONS**
All sentences with “date” and “time” NEs are automatically weighted as candidates.
**PRECISION:** Only “date” and “time” NEs from best candidate sentence are returned.

**HOW QUESTIONS**
If (question contains: “How much/many/long/…/often/tall/old…”):
\[ \rightarrow \text{Sentences with numbers (digits or words) are weighted as candidates.} \]
\[ \rightarrow \text{PRECISION: Only number entity from best candidate sentence is returned.} \]
Else: Use Enhanced Overlap Approach

**Expansion/Question as Set**
- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

**Final Answer**
**Increase Precision**
Remove question words from answer

**Best Answer**
ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate using:
• Pure overlap
• NP Overlap
• VP Overlap
*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.
For Who and Where questions increase PRECISION; Return only “person,” “organization,”
and “location” NEs from best answer.

WHEN QUESTIONS
All sentences with “date” and “time” NEs are automatically weighted as candidates.
PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS
If (question contains: “How much/many/long/.../often/tall/old...”):
→ Sentences with numbers (digits or words) are weighted as candidates.
→ PRECISION: Only number entity from best candidate sentence is returned.
Else: Use Enhanced Overlap Approach

Expansion/Question as Set
- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Input Files

Preprocessing

Story {Story ID, {Questions}}

Initialize NLP Tools

Parse Data

Processing

Question {Question ID, Question Type}

Question {Question ID, Question Type}

Question {Question ID, Question Type}

ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate using:
• Pure overlap
• NP Overlap
• VP Overlap
*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.
For Who and Where questions increase PRECISION; Return only “person,” “organization,”
and “location” NEs from best answer.

WHEN QUESTIONS
All sentences with “date” and “time” NEs are automatically weighted as candidates.
PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS
If (question contains: “How much/many/long/.../often/tall/old...”):
→ Sentences with numbers (digits or words) are weighted as candidates.
→ PRECISION: Only number entity from best candidate sentence is returned.
Else: Use Enhanced Overlap Approach

Expansion/Question as Set
- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Final Answer
Input Files

**Story**

{Story ID, {Questions}}

**Question**

{Question ID, Question Type}

...}

Parse Data

Initialize NLP Tools

Preprocessing

ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

**FIND CANDIDATES** – Weight each sentence as potential candidate using:

- Pure overlap
- NP Overlap
- VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.

For Who and Where questions increase PRECISION; Return only “person,” “organization,” and “location” NEs from best answer.

WHEN QUESTIONS

All sentences with “date” and “time” NEs are automatically weighted as candidates.

PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS

If (question contains: “How much/many/long/…/often/tall/old…”):

→ Sentences with numbers (digits or words) are weighted as candidates.

→ PRECISION: Only number entity from best candidate sentence is returned.

Else: Use Enhanced Overlap Approach

Final Answer

Increase Precision

Remove question words from answer

Best Answer

Expansion/Question as Set

- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Question Classification

Who

What

Where

Why

Which

Rest

When

How
private final static double littleBoost = 0.5;
private final static double defaultWeight = 1;
private final static double somewhatConfident = 1.5;
private final static double confident = 2;
private final static double veryConfident = 2.5;
private final static double extremelyConfident = 3.5;
private final static String filepaths = ".//src/";

Used levels of “Confidence” to weight different factors
- Overlap between question and sentence
- Existence of certain named entities
- Etc...
- Used for question types: \{Who, What, Where, Why, Rest\}
- Used for question types: {Who, What, Where, Why, Rest}
- Adds weight to candidate sentences for:
  - Pure overlap
  - NP overlap
  - VP overlap

```
private static String enhancedOverlapAnswer
(Question question, String[] sentences, double NPWeight, double VPWeight)
```
ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate using:
• Pure overlap
• NP Overlap
• VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.

For Who and Where questions increase PRECISION: Return only “person,” “organization,” and “location” NEs from best answer.

WHEN QUESTIONS

All sentences with “date” and “time” NEs are automatically weighted as candidates.

PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS

If (question contains: “How much/many/long/…/often/tall/old…”):
→ Sentences with numbers (digits or words) are weighted as candidates.
→ PRECISION: Only number entity from best candidate sentence is returned.

Else: Use Enhanced Overlap Approach

Expansion/Question as Set
- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Input Files

Story (Story ID, (Questions))

Question (Question ID, Question Type)

Question (Question ID, Question Type)

Initialize NLP Tools

Parse Data

Processing

Final Answer

Increase Precision

Remove question words from answer

Best Answer
Input Files

Preprocessing

Story (Story ID, [Questions])

Question (Question ID, Question Type)

Question (Question ID, Question Type)

Question (Question ID, Question Type)

... Initialize NLP Tools

Parse Data

Processing

ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate:

• Pure overlap
• NP Overlap
• VP Overlap

*Number of weight "points" awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.

For Who and Where questions increase PRECISION; Return only "person," "organization," and "location" NEs from best answer.

HOW QUESTIONS

If (question contains: "How much/many/long/.../often/tall/old..."):

→ Sentences with numbers (digits or words) are weighted as candidates.

→ PRECISION: Only number entity from best candidate sentence is returned.

Else: Use Enhanced Overlap Approach

WHEN QUESTIONS

All sentences with "date" and "time" NEs are automatically weighted as candidates.

PRECISION: Only "date" and "time" NEs from best candidate sentence are returned.

Expansion/Question as Set

- to lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

Final Answer

Best Answer

Increase Precision

Remove question words from answer
ENHANCED OVERLAP APPROACH

enhancedOverlap(Question question, sentences, NPWeight, VPWeight)*

FIND CANDIDATES – Weight each sentence as potential candidate using:
- Pure overlap
- NP Overlap
- VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.

ASSIGN BEST ANSWER to be sentence candidate with highest weight.

For Who and Where questions increase PRECISION; Return only “person,” “organization,” and “location” NEs from best answer.

WHEN QUESTIONS

All sentences with “date” and “time” NEs are automatically weighted as candidates.

PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

HOW QUESTIONS

If (question contains: “How much/many/long/…/often/tall/old…”):
→ Sentences with numbers (digits or words) are weighted as candidates.
→ PRECISION: Only number entity from best candidate sentence is returned.

Else: Use Enhanced Overlap Approach

[ H|h ]ow ( much | many | long | high | big | large | often | far | tall | old )

→ Look for number.
**Preprocessing**

- **Input Files**
  - **Story** (Story ID, [Questions])
  - **Question** (Question ID, Question Type)
  - **Question** (Question ID, Question Type)
  - **Question** (Question ID, Question Type)

- **Initialize NLP Tools**

**Processing**

- **ENHANCED OVERLAP APPROACH**
  - `enhancedOverlap(Question question, sentences, NPWeight, VPWeight)`

  **FIND CANDIDATES**
  - Weight each sentence as potential candidate using:
    - Pure overlap
    - NP Overlap
    - VP Overlap
  - *Number of weight “points” awarded for each type of overlap depend on question type.

- **ASSIGN BEST ANSWER** to be sentence candidate with highest weight.

  - **Expansion/Question as Set**
    - to lower case
    - Stemming
    - Synonyms of verbs
    - Remove closed class words

  - **WHEN QUESTIONS**
    - All sentences with “date” and “time” NEs are automatically weighted as candidates.
    - PRECISION: Only “date” and “time” NEs from best candidate sentence are returned.

  - **HOW QUESTIONS**
    - If (question contains: “How much/many/long/…/often/tall/old…”):
      - Sentences with numbers (digits or words) are weighted as candidates.
      - PRECISION: Only number entity from best candidate sentence is returned.
    - Else: Use Enhanced Overlap Approach

**Final Answer**

- **Increase Precision**
  - Remove question words from answer

**Best Answer**
## NLP Tools

<table>
<thead>
<tr>
<th>OpenNLP</th>
<th>WordNet with JAWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Named Entity Recognition</td>
<td>Noun/Verb Synonyms</td>
</tr>
<tr>
<td>POS Tagging</td>
<td></td>
</tr>
<tr>
<td>Stemming</td>
<td></td>
</tr>
<tr>
<td>Chunking</td>
<td></td>
</tr>
<tr>
<td>Sentence Segmentation</td>
<td></td>
</tr>
<tr>
<td>Tokenization</td>
<td></td>
</tr>
</tbody>
</table>
How Our System Works

What Did & Didn’t Work Well
Classify Questions
- who (people, places, org.)
- what ("what is")
- when (date or time period)
- where (geographic or everyday place - "kitchen")
- why
- how - "how much", "how many", "how can/did"

Other Questions

Text: Story

Question Tagging

Question Classification

Get relevant section of story

Tagging

Relevant excerpt

Tagging

Reformulation

Tagged excerpt

Answer Identification

Refinement

Synonyms + (verbal)

Morphology

Past/Present verb forms

NER # Pathways

KEY OBJECTIVE: 1. Translate question to things.
2. Translate story to things.
3. Identify answer using (1 + 2)
Less is more.

Occam’s Razor: Among competing hypotheses, the one with the fewest assumptions should be selected.

Used overlap as foundation and enhanced based on question type.
What didn’t work?

- Using machine learning for question classification
- Using coreferencing (OpenNLPs coreferencer wasn’t very accurate)
- NP synonyms (all over the place!)
- Using multiple candidate sentences to form answer
- ... etc;

Anything that we did which “overused” the question.
Given more time, we would...

- Use multiple sentences and coreference resolution
- Try alternatives to OpenNLP
- Incorporate WordNet hypernyms and hyponyms
- Find a way to use machine learning effectively
- Explore semantic class induction
**Preprocessing**

- **Story** {Story ID, {Questions}}
  - **Question** {Question ID, Question Type}
  - ... 
  - **Question** {Question ID, Question Type}

**Parse Data**

**Processing**

**ENHANCED OVERLAP APPROACH**

\[ \text{enhancedOverlap}(\text{Question, question, sentences, NPWeight, VPWeight}) \]

**FIND CANDIDATES** – Weight each sentence as potential candidate using:

- Pure overlap
- NP Overlap
- VP Overlap

*Number of weight “points” awarded for each type of overlap depend on question type.*

**ASSIGN BEST ANSWER** to be sentence candidate with highest weight.

For **Who** and **Where** questions increase **PRECISION**; Return only “person,” “organization,” and “location” NEs from best answer.

**WHEN QUESTIONS**

All sentences with “date” and “time” NEs are automatically weighted as **candidates**.

**PRECISION:** Only “date” and “time” NEs from best candidate sentence are returned.

**HOW QUESTIONS**

If (question contains: “How much/many/long/.../often/tall/old...”):

- Sentences with numbers (digits or words) are weighted as **candidates**.
- **PRECISION:** Only number entity from best candidate sentence is returned.

Else: Use **Enhanced Overlap Approach**

**Expansion/Question as Set**

- To lower case
- Stemming
- Synonyms of verbs
- Remove closed class words

**Question Classification**

- **Who**
- **What**
- **Where**
- **Why**
- **Which**
- **Rest**
- **When**
- **How**

**Final Answer**

**Increase Precision**

Remove question words from answer

**Best Answer**