Information Extraction over Structured Data: Question Answering with Freebase

Xuchen Yao and Benjamin Van Durme
“Who played in Gravity?”

- Bing: Satori

- Google: knowledge graph, Freebase
Answering from a Knowledge Base

Performances /film/film/starring

<table>
<thead>
<tr>
<th>Actor</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>George Clooney</td>
<td>Matt</td>
</tr>
<tr>
<td>Sandra Bullock</td>
<td>Ryan</td>
</tr>
</tbody>
</table>

- the model challenge
- the data challenge
QA from KB

The Model Challenge
Previous Approach: Semantic Parsing
Previous Approach: Semantic Parsing

Syntax:
- `city`
- `populous`
- `most`
- `in`
- `California`

Dep and CCG:

<table>
<thead>
<tr>
<th>What</th>
<th>states</th>
<th>border</th>
<th>Texas</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/ (S\NP)</td>
<td>N</td>
<td>(S\NP)/NP</td>
<td>NP</td>
</tr>
<tr>
<td>λf.λg.λx. f(x)∧g(x)</td>
<td>λx.state(x)</td>
<td>λx.λy.borders(y,x)</td>
<td>texaS</td>
</tr>
<tr>
<td>S/ (S\NP)</td>
<td>λg.λx.state(x)∧g(x)</td>
<td>S\NP</td>
<td>λy.borders(y,texas)</td>
</tr>
<tr>
<td>S</td>
<td>λx.state(x)∧ borders(x,texas)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Previous Approach: Semantic Parsing

Syntax

```
City

populous

in

most California
```

Semantics

```
\text{argmax}(\lambda x.\text{city}(x) \land \text{loc}(x, \text{CA}), \lambda x.\text{population}(x))
```

```
\text{what}
\begin{tabular}{ll}
\text{S/(S/\text{NP})/N} & \text{S/(S/\text{NP})} \\
\lambda f.\lambda g.\lambda x. f(x) \land g(x) & \lambda g.\lambda x.\text{state}(x) \\
\end{tabular}

\text{states}
\begin{tabular}{ll}
N & \text{border} \\
\lambda x.\text{state}(x) & \lambda x.\lambda y.\text{borders}(y, x) \\
\end{tabular}

\text{Texas}
\begin{tabular}{ll}
\text{S/\text{NP}} & \text{NP} \\
\text{texas} & \text{texas} \\
\end{tabular}

\text{first-order}
```
Previous Approach: Semantic Parsing

Syntax

\[
\text{city}
\]
\[
\text{populous}
\]
\[
\text{in}
\]
\[
\text{most}
\]
\[
\text{California}
\]

Semantics

\[
\text{argmax}(\lambda x.\text{city}(x) \land \text{loc}(x, \text{CA}), \lambda x.\text{population}(x))
\]

\begin{align*}
\text{What} & \quad \text{states} & \quad \text{border} & \quad \text{Texas} \\
\text{S/ (S\text{NP})/N} & \quad \lambda f. \lambda g. \lambda x. f(x) \land g(x) & \quad N & \quad \lambda x.\text{state}(x) \\
\text{NP} & \quad \text{NP} & \quad \text{NP} & \quad \text{NP} \\
\lambda g. \lambda x.\text{state}(x) \land g(x) & \quad \lambda x.\text{state}(x) \land \text{borders}(x, \text{texas}) & \quad \lambda y.\text{borders}(y, \text{texas}) & \quad \lambda y.\text{borders}(y, \text{texas}) \\
\text{S/ (S\text{NP})} & \quad \text{S/ (S\text{NP})} & \quad \text{S/ (S\text{NP})} & \quad \text{S/ (S\text{NP})} \\
\lambda g. \lambda x.\text{state}(x) \land g(x) & \quad \lambda g. \lambda x.\text{state}(x) \land g(x) & \quad \lambda g. \lambda x.\text{state}(x) \land g(x) & \quad \lambda g. \lambda x.\text{state}(x) \land g(x) \\
\text{S} & \quad \text{S} & \quad \text{S} & \quad \text{S} \\
\lambda x.\text{state}(x) \land \text{borders}(x, \text{texas}) & \quad \lambda x.\text{state}(x) \land \text{borders}(x, \text{texas}) & \quad \lambda x.\text{state}(x) \land \text{borders}(x, \text{texas}) & \quad \lambda x.\text{state}(x) \land \text{borders}(x, \text{texas})
\end{align*}
Previous Approach: Semantic Parsing

Syntax:
- city
- populous
- in
- most
- California

Semantics:
\[ \text{argmax}(\lambda x. \text{city}(x) \land \text{loc}(x, \text{CA}), \lambda x. \text{population}(x)) \]

Logic:
- \( \lambda \text{-DCS} \)

Parsing:
- dep
- ccg
- first-order

Queries:
- SPARQL
- SQL
- MQL

Table:
<table>
<thead>
<tr>
<th>What</th>
<th>states</th>
<th>border</th>
<th>Texas</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/ (S\NP)/N</td>
<td>N</td>
<td>(S\NP)/NP</td>
<td>NP</td>
</tr>
<tr>
<td>( \lambda f. \lambda g. \lambda x. f(x) \land g(x) )</td>
<td>( \lambda x. \text{state}(x) )</td>
<td>( \lambda x. \text{borders}(y, x) )</td>
<td>( \text{texas} )</td>
</tr>
<tr>
<td>S/ (S\NP)</td>
<td>S\NP</td>
<td>S\NP</td>
<td>S\NP</td>
</tr>
<tr>
<td>( \lambda g. \lambda x. \text{state}(x) \land g(x) )</td>
<td>( \lambda y. \text{borders}(y, \text{texas}) )</td>
<td>( \lambda x. \text{state}(x) \land \text{borders}(x, \text{texas}) )</td>
<td></td>
</tr>
</tbody>
</table>
Previous Approach: Semantic Parsing
Is this how YOU find the answer?

Syntax

city

populous

in

most

California

Semantics

argmax(\(\lambda x.\)city(\(x\)) \land \text{loc}(\(x\), CA), \(\lambda x.\)population(\(x\)))

\(\lambda\)-DCS

dep

ccg

parsing

logic

queries

What

states

border

Texas

\(S/ (S \backslash NP)/N\)

\(\lambda f.\lambda g.\lambda x. f(x) \land g(x)\)

N

\(\lambda x.\)state(\(x\))

\(S/ (S \backslash NP)\)

\(\lambda g.\lambda x.\)state(\(x\)) \land g(\(x\))

\((S \backslash NP)/NP\)

\(\lambda x.\lambda y.\)borders(\(y, x\))

NP

texas

\(S/ (S \backslash NP)\)

\(\lambda y.\)borders(\(y,\)texas)

\(S\backslash NP\)

\(\lambda y.\)borders(\(y,\)texas)

\(\lambda x.\)state(\(x\)) \land borders(\(x,\)texas)

SPARQL

SQL

MQL...
this instead might be how you find the answer

Question:

Who is the brother of Justin Bieber?
who is the brother of Justin Bieber?

1st step: go to JB’s Freebase page
**Who is the brother of Justin Bieber?**

2nd step: maybe wander around a bit?

<table>
<thead>
<tr>
<th>Celebrity</th>
<th>Relationship type</th>
<th>Start date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caitlin Beadles</td>
<td>Dated</td>
<td>2007</td>
<td>2007</td>
</tr>
<tr>
<td>Selena Gomez</td>
<td>Dated</td>
<td>2011 February</td>
<td>-</td>
</tr>
</tbody>
</table>

**Sexual orientation**

<table>
<thead>
<tr>
<th>Sexual orientation</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
</table>

**Legal entanglements**

<table>
<thead>
<tr>
<th>Offense</th>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driving under the influence</td>
<td>Miami</td>
<td>1/23/2014</td>
</tr>
<tr>
<td>Driving under the influence</td>
<td>Miami Beach</td>
<td>1/23/2014</td>
</tr>
</tbody>
</table>

**Substance abuse problems**
who is the brother of Justin Bieber?
finally: oh yeah, his brother
Freebase Topic Graph

we know **just enough** about the answer from the following view:
who is the brother of Justin Bieber?
Signals!

Major challenge for Question Answering:

finding indicative (linguistic) signals for answers
who is the brother of Justin Bieber?
QA on Freebase is now a binary classification problem on each node.

Features on Graph
extract features for each node

Justin Bieber
- has:awards_won
- has:place_of_birth
- has:sibling
- type:person
- ...

Jazmyn Bieber
- has:sibling
- gender:female
- type:person
- ...

Jaxon Bieber
- has:sibling
- gender:male
- type:person
- ...

brown: relation; relations connect to other nodes
blue: property; properties have literal values.
What do we know about the question?
What do we know about the question?
Features on Question

for every edge $e(s,t)$, extract: $s$, $t$, $s|t$, and $s|e|t$

- qword=what
- qfocus=name
- qverb=be
- qtopic=person
- qword=what|cop|qverb=be
- qword=what|nsubj|
- qfocus=name
- brother|nn|qtopic=person
- …
Combining Graph Features with Question Features

on graph

- Justin Bieber
  - has:awards_won
  - has:place_of_birth
  - has:sibling
  - type:person
  - ...

- Jazmyn Bieber
  - has:sibling
  - gender:female
  - type:person
  - ...

- Jaxon Bieber
  - has:sibling
  - gender:male
  - type:person
  - ...

on question

features

for every edge e(s,t), extract:

- s, t, s|t, and s|e|t

- qword=what
- qfocus=name
- qverb=be
- qtopic=person
- qword=what|cop|qverb=be
- qword=what|nsubj|qfocus=name
- brother|nn|qtopic=person
- ...

25
Some Combined Features Can Be Helpful

<table>
<thead>
<tr>
<th>Justin Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:awards_won</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:place_of_birth</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:sibling</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• type:person</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jazmyn Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender:female</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>brother</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jaxon Bieber (is answer)</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender: male</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>qword=what</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

brown: relation; relations connect to other nodes
blue: property; properties have literal values.
red: question features.
Some Combined Features Can Be Helpful

<table>
<thead>
<tr>
<th>Justin Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:awards_won</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:place_of_birth</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:sibling</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• type:person</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• ...</td>
<td>• ...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jazmyn Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender:female</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>brother</td>
</tr>
<tr>
<td>• ...</td>
<td>• ...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jaxon Bieber (is answer)</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender:male</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>qword=what</td>
</tr>
<tr>
<td>• ...</td>
<td>• ...</td>
</tr>
</tbody>
</table>

brown: relation; relations connect to other nodes
blue: property; properties have literal values.
red: question features.
Some Combined Features Can Be Helpful

<table>
<thead>
<tr>
<th>Justin Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:awards_won</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:place_of_birth</td>
<td>qword=what</td>
</tr>
<tr>
<td>• has:sibling</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• type:person</td>
<td>qfocus=name</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jazmyn Bieber</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender:female</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>brother</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jaxon Bieber (is answer)</th>
<th>expected weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• has:sibling</td>
<td>brother</td>
</tr>
<tr>
<td>• gender:male</td>
<td>brother</td>
</tr>
<tr>
<td>• type:person</td>
<td>qword=what</td>
</tr>
<tr>
<td>• ...</td>
<td></td>
</tr>
</tbody>
</table>

brown: relation; relations connect to other nodes
blue: property; properties have literal values.
red: question features.
who is the brother of Justin Bieber?

Justin Bieber?  
Jaxon Bieber  
Jasmin Bieber?  
...

Q → simple queries → features → classify → A
Information Extraction vs. Semantic Parsing

Q → simple queries → features → classify → A

Q → parses → logic → structured queries → A
QA from KB

The Data Challenge
Freebase Topic Graph

who is the *brother* of Justin Bieber?

How does a computer know this mapping?
The Challenge
aligning KB relations with NL words

• KB entry:
  – film/starring (Gravity, Bullock/Clooney)

• How questions can be asked:
  – what's the cast of Gravity?
  – who played/acted in Gravity?
  – who starred in Gravity?
  – show me the actors in Gravity.
Aligning KB Relations with NL Words

• annotated ClueWeb (with Freebase entities), released by Google
  – Sandra then was cast in Gravity, a two actor spotlight film
  – Sandra Bullock plays an astronaut hurtling through space in new blockbuster "Gravity"
  – Sandra Bullock stars/acts in Gravity
  – Sandra Bullock conquered her fears to play the lead in Gravity
Aligning KB Relations with NL Words

- annotated ClueWeb (with Freebase entities), thanks to Google
  - Sandra then was cast in Gravity, a two actor spotlight film
  - Sandra Bullock plays an astronaut hurtling through space in new blockbuster "Gravity"
  - Sandra Bullock stars/acts in Gravity
  - Sandra Bullock conquered her fears to play the lead in Gravity
Aligning KB Relations with NL Words

• Input: **film/starring** *(Gravity, Sandra Bullock)*
  – **Sandra** then was cast in **Gravity**, a two actor spotlight film
  – **Sandra Bullock** plays an astronaut hurtling through space in new blockbuster "**Gravity**"
  – **Sandra Bullock** stars/acts in **Gravity**
  – **Sandra Bullock** conquered her fears to play the lead in **Gravity**

• Task: find NL words that express **film/starring**
Aligning KB Relations with NL Words

• Input: **film/starring** (Gravity, Sandra Bullock)
  – Sandra then **was cast in** Gravity, a two actor spotlight film
  – Sandra Bullock **plays** an astronaut hurtling through space in new blockbuster "Gravity"
  – Sandra Bullock **stars/acts in** Gravity
  – Sandra Bullock conquered her fears to **play** the lead in Gravity

• Task: find **NL words** that express **film/starring**
Aligning KB Relations with NL Words

• maps the **NL phrases** to KB relations **film/starring**:
  – Sandra then **was cast in Gravity**, a two actor spotlight film
  – Sandra Bullock **plays** an astronaut hurtling through space in new blockbuster "Gravity"
  – Sandra Bullock **stars/acts in Gravity**
  – Sandra Bullock conquered her fears to **play the lead in Gravity**

• in massive scale:
  – Freebase: 40 million entities, 2.5 billion facts
  – ClueWeb09 Annotation: 5 billion **entities** in 340 million documents (5TB compressed)

• very simple solution:
  – treat it as an alignment problem (IBM Model 1)
  – fire up GIZA++ and hundreds of computers
### Samples of CluewebMapping

<table>
<thead>
<tr>
<th>Category</th>
<th>Sample Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>film.actor</td>
<td>• won, star, among, show, …</td>
</tr>
<tr>
<td>film.directed_by</td>
<td>• director, direct, by, with, …</td>
</tr>
<tr>
<td>celebrity.infidelity.victim</td>
<td>• Jennifer Aniston…</td>
</tr>
<tr>
<td>celebrity.infidelity.participant</td>
<td>• you know who…</td>
</tr>
</tbody>
</table>
## Samples of CluewebMapping

<table>
<thead>
<tr>
<th>Concept</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>film.actor</td>
<td>won, star, among, show, ...</td>
</tr>
<tr>
<td>film.directed_by</td>
<td>director, direct, by, with, ...</td>
</tr>
<tr>
<td>celebrity.infidelity.victim</td>
<td>Jennifer Aniston...</td>
</tr>
<tr>
<td>celebrity.infidelity.participant</td>
<td>you know who...</td>
</tr>
</tbody>
</table>
## Samples of CluewebMapping

<table>
<thead>
<tr>
<th>Term</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>film.actor</td>
<td>• won, star, among, show, …</td>
</tr>
<tr>
<td>film.directed_by</td>
<td>• director, direct, by, with, …</td>
</tr>
<tr>
<td>celebrity.infidelity. victim</td>
<td>• Jennifer Aniston…</td>
</tr>
<tr>
<td>celebrity.infidelity. participant</td>
<td>• you know who…</td>
</tr>
</tbody>
</table>
Samples of CluewebMapping

- **film.actor**: won, star, among, show, …
- **film.directed_by**: director, direct, by, with, …
- **celebrity.infidelity.victim**: Jennifer Aniston…
- **celebrity.infidelity.participant**: you know who…
Using KB Alignment as Features

• Who is the brother of Justin Bieber?
• predictions from KB alignment:
  – /people/sibling_relationship/sibling
  – /fictional_universe/
    sibling_relationship_of_fictional_characters/siblings
  – ...

• Features: the rank (top 1/3/5/50…) of node’s relation predicted by KB alignment
Evaluation

- Data: WebQuestions
- Berant et. al. (2013)
- 5810 questions annotated from 1 million crawled off Google Suggest

![Google Suggest Example](image-url)
Evaluation

- Data: WebQuestions
- Berant et. al. (2013)
- 5810 questions annotated from crawling off Google Suggest

- which states does the Connecticut river flow through?
- who does David James play for 2011?
- what date was John Adams elected president?
- what kind of currency does Cuba use?
- who owns the Cleveland Browns?
Evaluation

- Data: WebQuestions
- Berant et. al. (2013)
- 5810 questions annotated from crawling off Google Suggest

- which states does the *Connecticut* river flow through?
- who does *David James* play for 2011?
- what date was *John Adams* elected president?
- what kind of currency does *Cuba* use?
- who owns the *Cleveland Browns*?
Evaluation

- Tag named entities with Stanford CoreNLP (caseless model)
- Search named entities using the Freebase Search API
- Retrieve topics using the Freebase Topic API
- which states does the Connecticut river flow through?
- who does David James play for 2011?
- what date was John Adams elected president?
- what kind of currency does Cuba use?
- who owns the Cleveland Browns?
Evaluation

- Tag named entities with Stanford CoreNLP (caseless model)
- Search named entities using the Freebase Search API
- Retrieve topics using the Freebase Topic API

who did natalie portman play in star wars?

<table>
<thead>
<tr>
<th>topic</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>natalie_portman</td>
<td>722</td>
</tr>
<tr>
<td>star_wars</td>
<td>233</td>
</tr>
<tr>
<td>saturday_night_live_season_31</td>
<td>56</td>
</tr>
<tr>
<td>clone_wars</td>
<td>51</td>
</tr>
<tr>
<td>lego_star_wars</td>
<td>38</td>
</tr>
<tr>
<td>star_wars_music</td>
<td>37</td>
</tr>
<tr>
<td>star_wars_episode_iv_a_new_hope</td>
<td>36</td>
</tr>
<tr>
<td>star_wars_episode_i_the_phantom_menace</td>
<td>35</td>
</tr>
</tbody>
</table>
Evaluation & Training

- Data: WebQuestions
- Berant et. al. (2013)
- 5810 questions annotated from crawling off Google Suggest

Training

- L1 regularized Logistic Regression with high-performance ML tool: Classias (Okazaki, 2009)
  - original: 7 million feature types extracted
  - training: 4 hours
  - after: 30 thousand features with non-zero weight
F1 on TEST (2032 questions)

with Gold Retrieval

with Freebase Search

Berant et. al. (2013)

<table>
<thead>
<tr>
<th>feature</th>
<th>weight</th>
<th>feature</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>qfocus=religion</td>
<td>type=Religion</td>
<td>8.60</td>
<td>qword=when</td>
</tr>
<tr>
<td>qfocus=money</td>
<td>type=Currency</td>
<td>5.56</td>
<td>qverb=border</td>
</tr>
<tr>
<td>qverb=die</td>
<td>type=CauseOfDeath</td>
<td>5.35</td>
<td>qverb=go</td>
</tr>
</tbody>
</table>
Free917?
(Cai and Yates 2013)

• jacana-freebase was not designed to handle \texttt{argmax()}, \texttt{count()} operators out of the box
  – how many companies are traded by the nyse
  – how many teams participate in the uefa
  – how many ships has nathanael herreshoff designed
  – in what year did motorola have the most revenue
Conclusion

• Model: an Information Extraction approach with massive features
  – a naïve baseline for semantic parsing based QA!

• Data: helps with mapping between NL words and KB relations
  – CluewebMapping: ~3000 Freebase relations <-> 10,000 words

• Wednesday 10:45–11:10 Semantics III
  – Semantic Parsing via Paraphrasing.
  – Jonathan Berant and Percy Liang

• Thursday, Semantic Parsing Workshop
  – Freebase QA: Information Extraction or Semantic Parsing?
  – Xuchen Yao, Jonathan Berant and Benjamin Van Durme
thank you
Error Analysis
## 20%~25%: Annotation Error

<table>
<thead>
<tr>
<th>Question</th>
<th>MTurk answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>what state does selena gomez?</td>
<td>New York City</td>
</tr>
<tr>
<td>how old is sacha baron cohen?</td>
<td>a URL</td>
</tr>
<tr>
<td>what two countries invaded poland in the beginning of ww2?</td>
<td>Germany</td>
</tr>
<tr>
<td>which countries border the us?</td>
<td>Canada</td>
</tr>
<tr>
<td>where is rome italy located on a map?</td>
<td>Rome</td>
</tr>
<tr>
<td>how much did adriana lima gain during pregnancy?</td>
<td>Spike Guys’ Choice Awards</td>
</tr>
<tr>
<td>what does thai mean?</td>
<td>Language</td>
</tr>
<tr>
<td>which wife did king henry behead?</td>
<td>Anne of the Thousand Days</td>
</tr>
<tr>
<td>what are the major cities in france?</td>
<td>Paris</td>
</tr>
<tr>
<td>what season did tony soprano get shot?</td>
<td>The Sopranos</td>
</tr>
</tbody>
</table>
15%~20%: “complicated” questions

- what did James K. Polk do before he was president?
- what is the Oregon Ducks 2012 football schedule?
- what country did Germany invade first in WW1?
- who is governor of Ohio 2011?
- when did Charles Goodyear invented rubber?
- who did France surrender to in WW2?
- who did George W. Bush run against for the second term?
- who was the leader of Soviet Union during WWII?
5%~10%: answer typing failure

• what things did martin luther king do?
• what town was martin luther king assassinated in?
• what electorate does anna bligh represent?
• what channel is the usa pageant on?
• what are some of the traditions of islam?
• what is the state flower of arizona?
• what did the islamic people believe in?
• what did the scientist chadwick discovered?
Other errors

• Freebase search error (10%)
• ill-formed web text (2% ~ 3%)