cocobean
A Socially Ranked Question Answering System

ethan deyoung | jerry ye | jimmy chen | Srinivasan Ramaswamy

advisor: marti hearst
introduction

- *Traditional* Keyword Search
- keywords treated independently
- no semantic meaning
- instant results
introduction

- Question Answering Systems
  - highly specific domain
  - hard problem

- human dependent
- latency
introduction

- Social Ranking Websites
  - users pick what is "good"
  - reorders results
introduction

- cocobean
  - simple NLP processing
  - instant results
  - users’ votes improve ordering
a quick example

- Ask

 cocobean | qa

who is the founder of microsoft?  answer
a quick example

Vote

William Gates was a basketball player and a subject of the 1994 documentary Hoop Dreams. Gates and Arthur Agee were teenagers from the Cabrini-Green housing projects of Chicago who hoped to hone their basketball talent into professional careers with the NBA. The documentary followed them through their recruitment by St. Joseph High School (Westchester, Illinois).
a quick example

---

**Improve**

* cocobean|qa

**who is the founder of microsoft?**

Results for question: *who is the founder of microsoft?*

11 results found in 3.0216 sec

1. **Software wars**

   Free software and proprietary software, including Microsoft and Apple Inc., is an example of the Evil Empire genre of humor that is endemic in hacker culture where Bill Gates, the founder of Microsoft, is represented in a way that is almost the exact opposite of the low-key, soft-spoken image he holds in the popular press. Gates is instead shown either as Palpatine, the ruler of the Galactic Empire (Star Wars), or evil Galactic Empire of the Star Wars.

2. **Tien-Min Li**

   Mao Zedong, Liu Shaoci, Deng Xiaoping, and Lin Biao Li had seven children. His second son and youngest child, Kai-Fu Lee, is a prominent computer science researcher, the founding president of Google China and previously the founder of Microsoft Research Asia. == Books == Chou En-Lai. Institute of International Relations, 1970. [Review of Chou En-lai. by Li Tien-Min]

3. **William Gates**

   For the founder of Microsoft, see Bill Gates. William Gates (b. ca. 1972) was a basketball player and a subject of the 1994 documentary Hoop Dreams. Gates and Arthur Agee were teenagers from the Cabrini-Green housing projects of Chicago who hoped to hone their basketball talent into professional careers with the NBA. The documentary followed them through their recruitment by St. Joseph High School (Westchester, Illinois)/St. Joseph High School
a quick example

- Improve

**cocobean|qa**

Who is the founder of Microsoft?

Results for question: who is the founder of Microsoft?

11 results found in 3.4338 sec

**William Gates**

William Gates was a basketball player and a subject of the 1994 documentary Hoop Dreams. Gates and Arthur Agee were teenagers from the Cabrini-Green housing projects of Chicago who hoped to hone their basketball talent into professional careers with the NBA. The documentary followed them through their recruitment by St. Joseph High School (Westchester, Illinois) and St. Joseph High School.

**Tien-Min Li**

Mao Zedong, Liu Shaoqi, Deng Xiaoping, and Lin Biao. Li had seven children. His second son and youngest child, Kai-Fu Lee, is a prominent computer science researcher, the founding president of Google China and previously the founder of Microsoft Research Asia. == Books == Chou En-Lai. Institute of International Relations, 1970 [Review of Chou En-lai, by Li Tien-Min]

**Instant messaging & messengers**

Reuters head of collaboration services, David Gurle, the founder of Microsoft's Real Time Communication and Collaboration business, surprised everybody by signing the first Inter-service provider connectivity agreement on September 2003. This historic agreement enabled AOL Instant Messenger (AIM, ICQ and MSN Messenger) users to talk with Reuters Messaging counterparts and vice-versa agains an access fee. Following this breakthrough agreement between networks Microsoft.
how we did it

- User Interface Design
- System Architecture
- Ranking
ui design
usability study

- Initial Interviews
- Personas

Jason
Lurker
Read but never contributes

Linda
Casual User
Participates and contributes occasionally

Lee
Hard Core User
Active Participant
usability study

- Initial Interviews
- Personas

Jason
Lurker
Read but never contributes

Linda
Casual User
Participates and contributes occasionally

Lee
Hard Core User
Active Participant
usability study

• Initial Interviews
• Personas

Jason

Lurker
Read but never contributes

Linda

Casual User
Participates and contributes occasionally

Lee

Hard Core User
Active Participant
usability study

- Initial Interviews
- Personas

Jason
Lurker
Read but never contributes

Linda
Casual User
Participates and contributes occasionally

Lee
Hard Core User
Active Participant
usability study

- Initial Interviews
- Personas

Primary Persona
- Jason
- Linda
- Lee

Lurker
- Read but never contributes

Casual User
- Participates and contributes occasionally

Hard Core User
- Active Participant
usability study

- Low-fi Testing
  - blocking
  - two voting interfaces
- Results
  - down arrow/remove confusing
  - thumbs up/down more natural
usability study

- before

introduction  ui design  system architecture  ranking
usability study

- after

![Cocobean QA example](image)

**Question:** What is the statue of liberty?

**Answer:**

- **She's a Rebel:**
  - about United States|America, as the song is dated on July 4th and the song makes sense if the word "she" is replaced with "America". Or, in another popular interpretation, she could be viewed as the statue of liberty considering the emphasis of the song on symbolic representation. Its also important to note that the song Holiday (Green Day song) Holiday has a direct link to this song through the lyric dawning of our lives, which implies a strong skeptic view of US politics as it stands.

- **Punkrocker:**
  - of New Yorks skyline at nighttime. Everyone (except Iggy Pop who is seen in a car) head is replaced with a bear (so is the Statue of Liberty seen at the end.

- **League for Programming Freedom:**
  - thumbnail|Logo of the LPF League for Programming Freedom (LPF) was founded in 1989 by Richard Stallman to unite free software developers as well as developers of proprietary software to fight against software patents and the extension of the scope of copyright. Their logo is the Statue of Liberty holding a floppy disk and tape spool. The single event that had the most influence on the creation of the League was Apple Computer|Apples lawsuits against...
system architecture
system architecture

- components
  - wikipedia corpus
  - lucene search engine
- question transformation
- question classification
- ranking
- rendering
system architecture
Who is the president of France?
Who is the president of France?

Question Transformation Module

Question Classification Module
system architecture

Who is the president of France?

Question Transformation Module

Question Classification Module
Who is the president of France?

The president of France is ____, the president of France

Lucene

Wikipedia
Who is the president of France?

Question Transformation Module

The president of France is ___, the president of France.

Lucene

Wikipedia

Question Classification Module
Who is the president of France?

Question

Classification Module

Question Transformation Module

The president of France is ___, the president of France

Lucene

___ is the president of France

Results: IR Score, title, snippet

Wikipedia

Re-rank Module
system architecture

Who is the president of France?

The president of France is ___, the president of France

Lucene

Results: IR Score, title, snippet

class (Human)

Re-ranking Module

Wikipedia

Question Transformation Module

Question Classification Module
Who is the president of France?

Question Transformation Module

The president of France is ____, the president of France

Lucene

Re-ranking Module

Results: IR Score, title, snippet

class (Human)

fetch data: votes, reputation

Question Classification Module

Ranking Database
Who is the president of France?

The president of France is ____, the president of France

Lucene

Results: IR Score, title, snippet

class (Human)

Re-ranking Module

fetch data: votes, reputation

Ranking Database

Question Classification Module

Question Transformation Module

Who is the president of France?

Wikipedia
system architecture

Who is the president of France?

The president of France is ____, the president of France

Lucene

Results: IR Score, title, snippet

class (Human)

reordered results

fetch data: votes, reputation

Wikipedia
Who is the president of France?

Question Transformation Module

Lucene

The president of France is ___

Re-ranking Module

Results: IR Score, title, snippet

Wikipedia

fetch data: votes, reputation

reordered results

class (Human)

Rendering Module

Ranking Database

Question Classification Module

Results: IR Score, title, snippet
system architecture

Who is the president of France?

The president of France is ___, the president of France

Lucene

Results: IR Score, title, snippet

class (Human)

reordered results

fetch data: votes, reputation

cleaned and highlighted results

question classification module

question transformation module

rendering module

database

re-ranking module

wiki
Who is the president of France?

The president of France is ____. The president of France.

Lucene

Results: IR Score, title, snippet

class (Human)

Re-ranking Module

fetch data: votes, reputation

rendered results

cleaned and highlighted results

user votes

Ranking Database

Rendering Module

Question Transformation Module

Question Classification Module

Who is the president of France?
system architecture

Who is the president of France?

The president of France is ____, the president of France

Lucene

Results: IR Score, title, snippet

Re-ranking Module

class (Human)

cleaned and highlighted results

user votes

updated votes displayed to user

Question Transformation Module

Question Classification Module

Rendering Module

Ranking Database

Wikipedia

fetch data: votes, reputation

reordered results

Results: IR Score, title, snippet

Classification Module

Ranking

introduction

ui design

system architecture

ranking
ranking
features

- Images
- Votes Up
- Votes Down
- User Reputation
- Lucene IR Relevancy
- Question Type = Answer Type
- Rule Type
data set

- Gold Standard
  - Relevancy score for each question/answer pair
- Mechanical Turk
  - Pay people to vote
training

- Learn patterns from data
- Learn importance of each feature
- Support Vector Machine for Regression
**evaluation**

\[ NDCG = Z \sum_{i=1}^{k} \frac{score(i)}{\log_2(i + 1)} \]

<table>
<thead>
<tr>
<th></th>
<th>NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>0.63</td>
</tr>
<tr>
<td>random</td>
<td>0.65</td>
</tr>
<tr>
<td>no votes</td>
<td>0.79</td>
</tr>
<tr>
<td>with votes</td>
<td>0.91</td>
</tr>
<tr>
<td>best</td>
<td>1</td>
</tr>
</tbody>
</table>

- Normalized Discounted Cumulative Gain
- Document ranking ordered by SVM
- `score()` provided by golden set
- DCG improved in tests
results

- What Works
- What Doesn’t
questions

- Answers