Query Clarification in Voice Search
The case of ambiguous terms and false memories

Master’s Thesis by Arefeh Bahrami

Referees:
Prof. Dr. Benno Stein
Prof. Dr. Matthias Hagen

Advisor:
Johannes Kiesel
Clarification in text-based search

Searches related to who is heisenberg

- heisenberg principle
- heisenberg breaking bad
- werner heisenberg
- werner heisenberg experiment
- elisabeth heisenberg
Clarification in text-based search

Google search for "who is heisenberg"

Werner Heisenberg was a German physicist and philosopher who is noted for his crucial contributions to quantum mechanics. He devised a method to formulate quantum mechanics in terms of matrices, for which he was awarded the 1932 Nobel Prize for Physics.

Werner Heisenberg - Biography, Facts and Pictures - Famous Scientists
https://www.famousscientists.org/werner-heisenberg/

People also ask
- What is Heisenberg most famous for?
- Why did Walt pick the name Heisenberg?
- Who is Heisenberg in Breaking Bad?
- What is the Heisenberg effect?
How about voice-based search?

HOW CAN I DO QUERY CLARIFICATION

IN VOICE SEARCH?
The Scenario

What year is “Life on Mars” produced?
What year is “Life on Mars” produced?
The Scenario

What year is “Life on Mars” produced?
The Scenario

What year is “Life on Mars” produced?
The Scenario

What year is “Life on Mars” produced?
What year is “Life on Mars” produced?
Motivation

The Scenario

What year is “Life on Mars” produced?

Arefeh Bahrami

Query Clarification in Voice Search

October 26, 2018
The Scenario

What year is “Life on Mars” produced?
The Scenario

What year is “Life on Mars” produced?

2008
The Scenario

What year is “Life on Mars” produced?

Did you mean the song from David Bowie, the US television series, or the Lego theme?
What year is “Life on Mars” produced?

Did you mean the song from David Bowie from the album Hunky Dory, the US television series, originally aired on ABC, or the Lego space theme released in 2001?
The Scenario

What year is “Life on Mars” produced?

Did you mean the song from David Bowie, the US television series, the Lego theme, the album by Dexter Wansel, the British television series, or the poetry book...
What was the first option?
Research questions

- **RQ I:** Does the user background affect user satisfaction when experiencing query clarification?

- **RQ II:** Do the length and the number of the clarifying options affect the user satisfaction?

- **RQ III:** What is the best way of clarifying the user’s intent?
Section 2

Study 1: Clarifying Ambiguous Queries
Study 1: Clarifying Ambiguous Queries

Response types & Topics

- Direct
- 3-meanings
- 3-categories
- Many meanings (5-meanings)
- Longer explanations (3-long-meanings)
- Verify
Response types & Topics

- Direct
- 3-meanings
- 3-categories
- Many meanings (5-meanings)
- Longer explanations (3-long-meanings)
- Verify

Webis-Ambient-15
Study setup

Create mock-up skill for Amazon Alexa
- Natural set-up
- Real-time answer
Study setup

Task example:

Scenario: You want to surprise your Irish partner with an Irish cocktail called B-52, but you don’t know how to make it.

Interaction start:

- Alexa. Find!
- How to do a B-52?

After interaction:

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>The system answered my question</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system behaved as I expected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system was easy to hear/understand</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system was pleasant to use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Study setup

Task example:

Scenario: You want to surprise your Irish partner with an Irish cocktail called B-52, but you don’t know how to make it.

Interaction start:
- Alexa. Find!
- How to do a B-52?

After interaction:

- The system answered my question
- The system behaved as I expected
- The system was easy to hear/understand
- The system was pleasant to use

<table>
<thead>
<tr>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 14 participants (9 proficient, 5 intermediate), 13 main tasks
- 182 interaction phases
- 728 ratings
Research questions

- **RQ I:** Does the user background affect user satisfaction when experiencing query clarification?

- **RQ II:** Do the length and the number of the clarifying options affect the user satisfaction?

- **RQ III:** Does the user satisfaction decrease when the system asks for clarification?
Results

Understandability and pleasantness by English fluency level

The system was easy to hear/understand

The system was pleasant to use

English proficiency
- intermediate
- proficient

Ratings density
- agree
- neutral
- disagree

Arefeh Bahrami
Query Clarification in Voice Search
October 26, 2018
Results

Predictability and understandability by frequency of a participant’s voice interface usage:

The system behaved as I expected

The system was easy to hear/understand
Research questions

- **RQ I:** Does the user background affect user satisfaction when experiencing query clarification?

- **RQ II:** Do the length and the number of the clarifying options affect the user satisfaction?

- **RQ III:** Does the user satisfaction decrease when the system asks for clarification?
Results

Distribution and mean (○) of pleasantness for participants with intermediate English level by response method:

The system was pleasant to use

3-meaning response method ($\mu_p = 1.4$)
Study 1: Clarifying Ambiguous Queries

Research questions

- **RQ I:** Does the user background affect user satisfaction when experiencing query clarification?

- **RQ II:** Do the length and the number of the clarifying options affect the user satisfaction?

- **RQ III:** Does the user satisfaction decrease when the system ask for clarification?
Results

Response-specific ratings for predictability and pleasantness:

The system behaved as I expected

The system was pleasant to use

Response methods
- direct (hit)
- verify (hit)
- verify (miss)
- direct (miss)
Section 3

Study 2: Clarifying Queries with False Memories
Known-item search

Re-finding previously seen/heard items is not that easy!

WHERE HAVE I

SEEN THIS BEFORE
False memories in question

What is this film with Maggie Smith?

Ok I watched this ages ago but forgot what its called, ok so there's this teacher and she has an affair with a student and gets into trouble and Maggie Smith plays a creepy woman trying to take care of her any ideas?

Update: Okay erm to add more, sorry I can't really remember.
Ok so she ends up hated by everyone because of her sleeping with the student, the student still loves her but everyone hates her except for Maggie Smith who loves her and the girl moves in with Maggie and then at the end finds out that Maggie loves her.

🌟 Follow ⭐ 3 answers

Best Answer: Are you positive it is Maggie Smith because what you described sounds like "Notes on a Scandal" but that stars Judi Dench: [http://www.imdb.com/title/tt0465551/](http://www.imdb.com/title/tt0465551/)
“Movie”..Maggie Smith plays a creepy woman...”
Study 2: Clarifying Queries with False Memories

Movie”..Maggie Smith plays a creepy woman…”

Question

Answer: "Judi Dench and Cate Blanchett star in "Notes on a Scandal"."
Hmm...

**How do I correct false memories?**
Research questions

- **RQ I:** Does language fluency affect user satisfaction?
- **RQ II:** Do wrong clarifications degrade user satisfaction?
- **RQ III:** How to best clarify false memories?
Response types & Topics

- **None** no clarification ("Sorry, I don’t know that one!")
- **Direct** answer query for a changed detail without informing users.
- **Negatively Clarified** ("Sorry, I don’t know that one, but...")
- **Positively Clarified** ("You probably mean...")
Response types & Topics

- **None** no clarification ("Sorry, I don’t know that one!")
- **Direct** answer query for a changed detail without informing users.
- **Negatively Clarified** ("Sorry, I don’t know that one, but...")
- **Positively Clarified** ("You probably mean...")

**Webis-KIQC-13**

annotations for 2,755 questions
Study setup

Task example:

Scenario: You try to remember the title of a controversial book that came out back in the 1990s and claimed scientific evidence that whites are genetically superior to blacks. You think it was called something like "The something Factor."

Interaction start: Alexa. Explore!
What is the title of the book from the 1990s that claimed superiority of Whites and is called "The something Factor"?

Post-interaction questions: Agree Neutral Disagree Don’t know
The system was helpful
The system behaved as I expected
The system was easy to hear/understand
The system was pleasant to use
Study setup

Task example:

Scenario: You try to remember the title of a controversial book that came out back in the 1990s and claimed scientific evidence that whites are genetically superior to blacks. You think it was called something like “The *something* Factor.”

Interaction start: Alexa. Explore!
What is the title of the book from the 1990s that claimed superiority of Whites and is called ”The *something* Factor”?

Post-interaction questions:
- The system was helpful
- The system behaved as I expected
- The system was easy to hear/understand
- The system was pleasant to use

- 12 participants (7 proficient, 5 intermediate), 14 main tasks
- 168 interaction phases
- 672 ratings
Research questions

- **RQ I:** Does language fluency affect user satisfaction?
- **RQ II:** Do wrong clarifications degrade user satisfaction?
- **RQ III:** How to best clarify false memories?
Results

Effectiveness and predictability by English fluency:

The system was helpful

The system behaved as I expected
Results

Understandability and pleasantness by English fluency

The system was easy to hear/understand

The system was pleasant to use
Research questions

- **RQ I:** Does language fluency affect user satisfaction?
- **RQ II:** Do wrong clarifications degrade user satisfaction?
- **RQ III:** How to best clarify false memories?
Results

Distribution and mean (○) of pleasantness by response method and changed detail:

The system was pleasant to use

The system was pleasant to use
Research questions

- **RQ I:** Does language fluency affect user satisfaction?
- **RQ II:** Do wrong clarifications degrade user satisfaction?
- **RQ III:** How to best clarify false memories?
Study 2: Clarifying Queries with False Memories

Results

Response-specific ratings for predictability and pleasantness when the false memory detail is corrected:

The system behaved as I expected

The system was pleasant to use
Summary

- First step towards the task of voice search query clarification.

- 1\textsuperscript{st} Study: clarifying ambiguous queries, 7 response types, mock-up skill for Amazon Alexa, results analysis.

- 2\textsuperscript{nd} Study: clarifying queries with false memories, 4 response types, mock-up skill, results analysis.
Conclusion & Future work

- Users won’t mind being asked for clarification.
- Consider English fluency level.
- Use positive tone in the response.
- Three clarification options work best.

- Test alternative response methods.
- Use crowdsourcing platforms.
Users won’t mind being asked for clarification.

Consider English fluency level.

Use positive tone in the response.

Three clarification options work best.

Test alternative response methods.

Use crowdsourcing platforms.

Thank you! 😊