From Session Detection to Mission Detection

Matthias Hagen Jakob Gomoll Anna Beyer Benno Stein

Bauhaus-Universität Weimar
matthias.hagen@uni-weimar.de

OAIR 2013
Lisbon, Portugal
May 23, 2013

What is the user searching?

`manhattan`

Without context . . .

source: [http://usatravel.about.com/od/Pictures-And-Maps/ss/Amazing-Aerial-Views-Of-America.htm]

What if you knew the previous queries?

```
party ideas
cocktail recipes
caipirinha
manhattan
```

What if you knew the previous queries?

```
party ideas
cocktail recipes
caipirinha
manhattan
```

source: [https://commons.wikimedia.org/wiki/File:Manhattan_Cocktail2.jpg]

What if you knew the previous queries?

```
party ideas
cocktail recipes
caipirinha
manhattan
```

Improves

- Intent understanding
- Retrieval precision
- Disambiguation

source: [https://commons.wikimedia.org/wiki/File:Manhattan_Cocktail2.jpg]

A typical query log

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Physical sessions (gaps $\leq$ 30 minutes)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Physical sessions → interleaved intents

Query	Time	Intent
ancient turkey	2013-04-20 20:02:44	
history istanbul	2013-04-20 20:24:17	
istanbul archeology	2013-04-21 12:02:54	
istanbul archeology	2013-04-21 18:31:21	history
weather new york	2013-04-21 18:45:23	weather
constantinople	2013-04-21 18:45:36	history
footbal lisbon	2013-04-21 19:14:01	sports
football lisbon	2013-04-21 19:14:11	sports
benfica vs sporting	2013-04-21 20:23:04	
derby eterno	2013-04-21 22:42:48	
constantinople	2013-04-21 23:09:02	
constantinople	2013-04-21 23:27:38	

Actual search intent switches

Query	Time	Intent
ancient turkey	2013-04-20 20:02:44	history
history istanbul	2013-04-20 20:24:17	
istanbul archeology	2013-04-21 12:02:54	
istanbul archeology	2013-04-21 18:31:21	
weather new york	2013-04-21 18:45:23	weather
constantinople	2013-04-21 18:45:36	history
footbal lisbon	2013-04-21 19:14:01	sports
football lisbon	2013-04-21 19:14:11	
benfica vs sporting	2013-04-21 20:23:04	
derby eterno	2013-04-21 22:42:48	
constantinople	2013-04-21 23:09:02	history
constantinople	2013-04-21 23:27:38	

Long-term tasks

Query	Time	Intent
ancient turkey	2013-04-20 20:02:44	history
history istanbul	2013-04-20 20:24:17	
istanbul archeology	2013-04-21 12:02:54	
istanbul archeology	2013-04-21 18:31:21	
weather new york	2013-04-21 18:45:23	weather
constantinople	2013-04-21 18:45:36	history
footbal lisbon	2013-04-21 19:14:01	sports
football lisbon	2013-04-21 19:14:11	
benfica vs sporting	2013-04-21 20:23:04	
derby eterno	2013-04-21 22:42:48	
constantinople	2013-04-21 23:09:02	history
constantinople	2013-04-21 23:27:38	

Multitasking and search missions

Observations [Spink et al., 2006; Jones and Klinkner, 2008]

Physical sessions:	interleaved intents	(multitasking)
Long-term tasks:	several sessions	(search missions)

Multitasking and search missions

Observations

[Spink et al., 2006; Jones and Klinkner, 2008]

Physical sessions:	interleaved intents	(multitasking)
Long-term tasks:	several sessions	(search missions)

Traditional session detection

Only consecutive queries $\rightarrow$ Missions impossible

Example

`history istanbul`	2013-04-20 20:24:17	same ✓
`istanbul archeology`	2013-04-21 12:02:54	
— — — — — — — — —		new ✓
`football lisbon`	2013-04-21 19:14:11	
— — — — — — — — —		new ↯
`constantinople`	2013-04-21 23:09:02	

Our topic ...

Pre-retrieval session + mission detection

Our topic ...

Pre-retrieval session + mission detection

Remark: Runtime is crucial!

Typical query similarity features

Temporal thresholds	5 minutes	[Silverstein et al., 1999]
	15 minutes	[He and Göker, 2000]
	30 minutes	[Downey et al., 2007]
	120 minutes	[Buzikashvili and Jansen, 2006]
	user specific	[Murray et al., 2006]
Lexical similarity	term overlap	[Kotov et al., 2011]
	n-gram overlap	[Zhang and Moffat, 2006]
	Levenshtein distance	[Jones and Klinkner, 2008]
	reformulation patterns	[Huang and Efthimiadis., 2009]
Semantic similarity	ESA	[Lucchese et al., 2011]
	Search results	[Radlinski and Joachims, 2005]
	Linked Open Data	[Hollink et al., 2011]

Previous methods

Feature combinations

- More accurate than single features
- One of the best: Geometric method (time + lexical) [Gayo-Avello, 2009]

Previous methods

Feature combinations

- More accurate than single features
- One of the best: Geometric method (time + lexical) [Gayo-Avello, 2009]

Shortcomings

- All features evaluated simultaneously $\rightarrow$ runtime
- Geometric method ignores semantics $\rightarrow$ accuracy

Examples

Substring test suffices.	Geometric method fails.
`football`	`benfica vs sporting`
`football lisbon`	`derby eterno`

Our previous cascading method . . . [Hagen et al., 2011]

source: [http://wp.ltchambon.com/wp-content/uploads/2010/09/Cascade-de-Tufs-Baume-les-messieurs-Jura.jpg]

. . . well . . . it looked more like this

[Hagen et al., 2011]

source: [http://www.solarshop.com/solarpix/Solar Cascade 4 Tier GreenL.jpg]

. . . well . . . it looked more like this

[Hagen et al., 2011]

source: [http://www.solarshop.com/solarpix/Solar Cascade 4 Tier GreenL.jpg]

Step 1: Subset test

↘

Step 2: Geometric method

↘

Step 3: ESA similarity

↘

Step 4: Search results

Basic idea

Increased feature cost (runtime) from step to step.
Expensive features only if previous steps "unreliable."

Our improved cascade

source: [http://www.solarshop.com/solarpix/Solar Cascade 4 Tier GreenL.jpg]

Basic idea still

Cheap features first.

Step 0: Time gaps

↘

Step 1: Substring test

↘

Step 2: Lexical similarity

↘

Step 3: ESA similarity

↙

Step 4: LOD similarity

↙

Step 5: Search results

Step 0: Time gaps

(gaps $\leq$ 90 minutes)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Step 1: Substring test (one query contained in the other)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Step 2: Lexical similarity (cosine similarity of char 3-/4-grams)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Step 3: ESA similarity [Gabrilovich and Markovitch, 2007]

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Step 4: LOD similarity (heaviest 2-step path in DBpedia)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Step 5: Search results (shared top-10 result)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Computed logical sessions

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Unidentified mission connections

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

Mission detection

Idea

Run the cascade twice:

1. Session detection on query level
2. Mission detection on session level

Example: You are here!

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
▶ constantinople	2013-04-21 23:09:02 ◀
constantinople	2013-04-21 23:27:38

Example: You are here! (ignore intermediate queries)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
▶ constantinople	2013-04-21 23:09:02 ◀
constantinople	2013-04-21 23:27:38

Step 1 of the cascade does the job (substring test)

Query	Time
ancient turkey	2013-04-20 20:02:44
history istanbul	2013-04-20 20:24:17
istanbul archeology	2013-04-21 12:02:54
istanbul archeology	2013-04-21 18:31:21
weather new york	2013-04-21 18:45:23
constantinople	2013-04-21 18:45:36
footbal lisbon	2013-04-21 19:14:01
football lisbon	2013-04-21 19:14:11
benfica vs sporting	2013-04-21 20:23:04
derby eterno	2013-04-21 22:42:48
constantinople	2013-04-21 23:09:02
constantinople	2013-04-21 23:27:38

How good does it work?

Accuracy

Runtime

Available evaluation corpora

Gayo-Avello's corpus (AOL log, 1 annotator)

- 11 500 queries — But: empty queries, order changed
- 215 users — But: many with ≤ 3 queries
- 2.7 queries per session — But: no mission annotation

Lucchese et al.'s corpus (AOL log, 1 annotator)

- 1500 queries — But: 97% of queries dropped
- 13 users

Our new corpus (basis: Gayo-Avello, 2 annotators)

- 8800 queries — Empty queries removed
- 127 users — Users with ≤ 3 queries removed
- 6.5 queries per mission

Available evaluation corpora

Gayo-Avello's corpus (AOL log, 1 annotator)

- 11 500 queries — But: empty queries, order changed
- 215 users — But: many with ≤ 3 queries
- 2.7 queries per session — But: no mission annotation

Lucchese et al.'s corpus (AOL log, 1 annotator)

- 1500 queries — But: 97% of queries dropped
- 13 users

Our new corpus (basis: Gayo-Avello, 2 annotators)

- 8800 queries — Empty queries removed
- 127 users — Users with ≤ 3 queries removed
- 6.5 queries per mission

Available evaluation corpora

Gayo-Avello's corpus (AOL log, 1 annotator)

- 11 500 queries
- 215 users
- 2.7 queries per session

But: empty queries, order changed

But: many with ≤ 3 queries

But: no mission annotation

Lucchese et al.'s corpus (AOL log, 1 annotator)

- 1500 queries
- 13 users

But: 97% of queries dropped

Our new corpus (basis: Gayo-Avello, 2 annotators)

- 8800 queries
- 127 users
- 6.5 queries per mission

Empty queries removed

Users with ≤ 3 queries removed

Session detection accuracy

F-Measure on our corpus (6630 queries, 25 % training)

Geometric method	0.821
Original cascade	0.853
Improved cascade	**0.946**

Session detection accuracy

F-Measure on our corpus (6630 queries, 25 % training)

Geometric method	0.821
Original cascade	0.853
Improved cascade	**0.946**

Performance per step

	decides	F-Measure	time
Step 0	23.87%	0.807	0.033 ms
Step 1	48.72%	0.845	0.002 ms
Step 2	13.28%	0.925	0.178 ms
Step 3	0.60%	0.930	0.237 ms
Step 4	0.11%	0.930	12.770 ms
Step 5	2.03%	0.946	13.359 ms

Remark: Without Steps 4–5 about 8500 queries per second!

Mission detection accuracy

Accuracy on our corpus (6630 queries, 25 % training)

- 865 continuations identified (269 missed: 157 horizon effects)
- 307 sessions wrongly assigned a continuation

Mission detection accuracy

Accuracy on our corpus (6630 queries, 25 % training)

- 865 continuations identified (269 missed: 157 horizon effects)
- 307 sessions wrongly assigned a continuation

Without semantic steps!

- 807 continuations identified
- 113 sessions wrongly assigned a continuation

Remark: Missions often picked up with identical query.

Mission detection accuracy

Accuracy on our corpus (6630 queries, 25 % training)

- 865 continuations identified (269 missed: 157 horizon effects)
- 307 sessions wrongly assigned a continuation

Without semantic steps!

- 807 continuations identified
- 113 sessions wrongly assigned a continuation

Remark: Missions often picked up with identical query.

Observations

- Session detection benefits from semantics
- Mission detection better without semantics

Almost the end: The take-home messages!

What we have done

Results

- Improved cascade
- Cheap features first
- Applicable to missions
- Semantics rather costly
- LOD not really useful yet

- Large mission corpus

Future Work

- Speed up semantics
- Prune LOD graph
- Wikipedia index
- Search results

What we have (not) done

Results

- Improved cascade
- Cheap features first
- Applicable to missions
- Semantics rather costly
- LOD not really useful yet

- Large mission corpus

Future Work

- Speed up semantics
- Prune LOD graph
- Wikipedia index
- Search results

What we have (not) done

Results

- Improved cascade
- Cheap features first
- Applicable to missions
- Semantics rather costly
- LOD not really useful yet

- Large mission corpus

Future Work

- Speed up semantics
- Prune LOD graph
- Wikipedia index
- Search results

Thank you

☺