Overview of NTCIR-9 1CLICK

Tetsuya Sakai
Makoto P. Kato
Young-In Song

December 9, 2011@NTCIR-9, Tokyo
TALK OUTLINE

• MOTIVATION
• TASK
• METRICS
• RESULTS
• TENTATIVE CONCLUSIONS
Traditional search vs One Click Access

Enter query

Click SEARCH button

Scan ranked list of URLs

Click URL

Read URL contents

Get all desired information

Enter query

Click SEARCH button

Get all desired information

Shonan Atsugi Hospital


Evaluate text based on nuggets
“1CLICK” Definition

• Given a query and an output window of size \( X \), return a text (“\( X \)-string”) that
  - Presents important nuggets first; and
  - Minimise the amount of text the user has to read (or the time spent)

• 1CLICK is direct: no user operation between clicking SEARCH button and seeing the output

• 1CLICK is immediate: very little time spent to locate relevant information within output
TALK OUTLINE

• MOTIVATION
• TASK
• METRICS
• RESULTS
• TENTATIVE CONCLUSIONS
1CLICK Queries

- 60 queries; 4 query types based on SIGIR’09 Good Abandonment paper
  - CELEBRITY date of birth, profession, hobbies, etc.
  - LOCATION contact info, address, opening hours, etc.
  - DEFINITION (of a term)
  - QA (Natural language questions)

- 2,839 nuggets
  - facts as of Dec 31, 2010

<table>
<thead>
<tr>
<th>Query type</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>126.5</td>
<td>[38, 368]</td>
</tr>
<tr>
<td>LO</td>
<td>30.9</td>
<td>[10, 125]</td>
</tr>
<tr>
<td>DE</td>
<td>26.3</td>
<td>[2, 174]</td>
</tr>
<tr>
<td>QA</td>
<td>5.6</td>
<td>[2, 26]</td>
</tr>
<tr>
<td>Total</td>
<td>47.3</td>
<td>[2, 368]</td>
</tr>
</tbody>
</table>

J. Li, S. Huffman, and A. Tokuda. Good abandonment in mobile and PC internet search. ACM SIGIR 2009
Apart from Ueno Zoo, where in Japan can I see a panda?

Which is taller, Tsutenkaku or Utsunomiya Tower?

The three duties of a Japanese citizen

What are the storage capacity units larger than MB, GB and TB?
1CLICK Nugget Record

Nugget = 5-tuple
- ID: N001
- Weight: 15
- Nugget semantics:
  - Osamu Tezuka was born on Nov 3, 1928.
- Vital string:
  - 1928.11.03
- Source URL:
  - http://tezukaosamu.net/jp/about/

An atomic factual statement in natural language. Used by nugget match assessors for identifying the nugget.

Used for approximating the minimal string length to convey the meaning of the nugget in X-string.
1CLICK Participants and Runs

- **DESKTOP** runs: $X=500$
  - (about five search engine snippets “above the fold”)
- **MOBILE** runs: $X=140$
  - (satisfy user with one mobile screen or a “tweet”)
- Participants: 25 teams registered, 3 submitted (10 runs)

Kyoto U (Information Extraction)
Tokyo Institute of Tech (Summarisation)
MSRA (Passage Retrieval)
TALK OUTLINE

• MOTIVATION
• TASK
• METRICS
• RESULTS
• TENTATIVE CONCLUSIONS
Analogy with nDCG

nDCG discounts documents based on ranks

S-measure discounts nuggets based on offsets (positions in X-string)

Ranked list of documents

X-string

Sakai, Kato, Song: Click the Search Button and Be Happy: Evaluating Direct and Immediate Information Access, ACM CIKM 2011
Offset-based linear discounting

Assumption: user’s reading speed is constant -> nugget value wears out linearly.

Position of a nugget match is taken as the END offset (user has to read THROUGH the nugget).
Manual nugget matching

Assessor selects a nugget match area and saves

Offset recorded

Content: 407 chars
2011年3月1日...公式サイト。、出演作品、ブロック。公式サイト。、、公式サイト...の売れ筋商品、出演DVD、写真集、CD、関連書籍など。http://www.amazon.co.jp/。Rosebud。ファンサイト。笑顔の写真集、事典、な...2008年7月14日...名前：（きたがわけいこ）出身地：兵庫県神戸市愛称：（けいちゃん、きたけい）生年月日：1986年8月22日、血液型：O型、身長：160cm体重：非公開所属事務所：スタダストプロモーション...2011年3月1日...--Yahoo!人物名鑑。2011年3月3日...の画像、動画、ニュース、作品情報、ブログ、Facebook（出典：Wikipedia）等、Yahoo!人物名鑑でチェック！http://talent.yahoo.co.jp/pf/detail/.../2010年1月8日...きたがわ・けいこ。ヒートアイランド。©2007「ヒートアイランド」製作委員会。木村了/城田優/。生年月日：1986/08/22。出身地：兵庫県。IMDBで検索[IMDBSearch]をYahoo

1CLICK encourages abstractive summarisation by paraphrasing etc. to save space -> automatic matching probably not sufficient
nDCG normalises system performance with ideal ranked list

Problem: Nugget lengths vary!

S-measure normalises system performance with PMO

Vital string: approximates a most concise nugget output

Sorted by weight and length
\[ S\text{-measure} = \frac{\sum_{m \in M} w(m) \max(0, L - \text{offset}(m))}{\sum_{n \in N} w(n) \max(0, L - \text{offset}^*(v(n)))} \]

- Present important nuggets first!
- Minimise the amount of text the user reads!
- Reduces to Weighted recall (W-recall) when \( L = \infty \)
- For Japanese, \( L = 500 \) characters \( \equiv \) one minute
TALK OUTLINE

• MOTIVATION
• TASK
• METRICS
• RESULTS
• TENTATIVE CONCLUSIONS
System ranking (Mean over 60 queries)

S-measure ($L=500$)

W-recall

Assessors A and B

Intersection

Union

a: KUIDL-D-OPEN-1
b: MSRA1click-D-OPEN-2
c: KUIDL-D-OPEN-2
d: MSRA1click-D-OPEN-1
e: KUIDL-M-OPEN-2
f: KUIDL-M-OPEN-1
g: TTOKU-D-ORCL-1
h: TTOKU-D-ORCL-2
i: TTOKU-M-ORCL-1
j: TTOKU-M-ORCL-2
Per-query disagreement: S vs W

Kobe city central library

Nuggets representing opening hours, address, phone/fax, access from nearest station etc.
**S vs readability vs trustworthiness**

High trustworthiness AND readability -> high S
Low trustworthiness AND readability -> low S
Nugget match evaluation effort

Average time to handle one X-string: 151 secs

CE queries with large lists of titles as nuggets are time-consuming.

Example:
- 0009 山崎邦正 (Hose Yamasaki): 92 nuggets; 358 seconds
- 0053 島田紳助 (Shinsuke Shimada): 225 nuggets; 432 seconds
- 0031 手塚治虫 (Osamu Tezuka): 368 nuggets; 425 seconds
- 0030 中川昭一 (Shoichi Nakagawa): 109 nuggets; 317 seconds
- 0010 浅田真央 (Mao Asada): 133 nuggets; 308 seconds
TALK OUTLINE

• MOTIVATION
• TASK
• METRICS
• RESULTS

• TENTATIVE CONCLUSIONS
Tentative Conclusions

• The 1CLICK evaluation framework is useful!
  – It can penalise systems that make the user read a lot of irrelevant text.

• The 1CLICK evaluation framework is feasible!
  – Average time to handle one X-string: 151 secs
  – The manual nugget match evaluation effort can be further reduced by carefully selecting queries and nugget types.
Japanese, English, Korean (,Chinese)

- Separate training data and test data (esp. Yahoo! Chiebukuro QA queries)
- Select realistic queries and nugget types (e.g. avoid long lists of comic book titles)
- Refine nugget definition
- Refine nugget match criteria
- Handle query ambiguity?
- More query types
- Penalizing irrelevant information
Query Types

**CELEBRITY** User wants to gather various facts about a celebrity: date/place of birth, real name, blood type, height, hobbies, profession, personal history, awards, publications, discography, films, TV series, favourite baseball team, favourite food etc.

**LOCAL** User wants to contact or visit a facility (school, shop, office, amusement park, hotel, train station etc.). Hence (s)he wants facts such as postal and email addresses, phone and fax numbers, opening hours, how to access the facility by train/bus/car, nearest stations, time required for the travel, whether the facility has a car park and its opening hours etc.

**DEFINITION** User seeks the definition of a term.

**QA** User seeks a short answer to a question.
Readability and Trustworthiness

• **Readability** is to do with coherence and cohesiveness, and how easy it is for the user to read and understand the text. For example, garbled text and the lack of spaces between two unrelated contexts can hurt readability.

• **Trustworthiness** means whether the user is likely to believe what it says in the X-string, as well as whether the user is likely to be misled. For example, if the X-string looks as if it was extracted from a source that is clearly not authoritative, it's not trustworthy. Moreover, if what is implied in the X-string is contrary to facts, it is not trustworthy.
Mean over 15 CE queries
Mean over 15 LO queries
Mean over 15 QA queries