Augmenting Web Search Surrogates with Images

Robert Capra\textsuperscript{1}
Jaime Arguello\textsuperscript{1}
Falk Scholer\textsuperscript{2}

\textsuperscript{1}School of Information and Library Science
University of North Carolina at Chapel Hill

\textsuperscript{2}RMIT University, Melbourne, Australia

CIKM 2013, San Francisco, CA
October 29, 2013
Web search results

Each result is represented on a search result page (SERP) by a *surrogate* consisting of a title, URL, and text snippet.
Other surrogates include images

- News articles
- Shopping results
- Videos
Why not include images for Web search results?

Could images help users assess relevance?
A particular case of interest...

Ambiguous query

Jaguar: Luxury Cars & Sports Cars | Jaguar USA
www.jaguarusa.com/

The official home of Jaguar USA. Our luxury cars feature innovative designs along with legendary performance to deliver one of the top sports cars in the ...
2014 Jaguar F-TYPE Price ... - 2013 Jaguar XF Price & Models - Models & Pricing
23,123 people +I'd this

Jaguar - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Jaguar

The jaguar is a big cat, a feline in the Panthera genus, and is the only Panthera species found in the Americas. The jaguar is the third-largest feline after the tiger ...

Jaguar Cars - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Jaguar_Cars

Jaguar Cars since December 2012 officially incorporated as Jaguar Land Rover Ltd. is a British multinational car manufacturer headquartered in Whitley ...

Jaguar Dealer in Raleigh NC - New Used Jaguar Cars Durham ...
www.leithjaguar.com/

Jaguar Dealer in Raleigh, NC. New Used & CPO Jaguar cars for sale. Jaguar Parts & service serving Raleigh Durham, Cary, Durham & Apex. Jaguar XF, XJ, XKR ...

Raleigh Used Jaguar Cars For Sale Cary Apex Durham NC
www.leithjaguar.com/usedused.aspx

Used & CPO Jaguar cars for sale in Raleigh NC. Used Jaguar dealer in Raleigh, Cary Apex, North Carolina. Used XJ, XKR, XF & F-Type cars for sale in inventory ...

Jaguars, Jaguar Pictures, Jaguar Facts - National Geographic
animals.nationalgeographic.com/animals/mammals/jaguar/

Learn all you wanted to know about Jaguars with pictures, videos, photos, facts, and news from National Geographic.

Jaguar News - Autoblog
www.autoblog.com > Automobiles

Read the latest Jaguar news and reviews complete with photos, videos, and road tests of new Jaguar vehicles.

Jacksonville Jaguars, Official Site of the Jacksonville Jaguars
www.jaguars.com/

The official team site with scores, news items, game schedule, and roster.
Diversified results

Ambiguous query

Diversified results
Diversified results

- Jaguar: Luxury Cars & Sports Cars | Jaguar USA
  www.jaguarusa.com
  The official home of Jaguar USA. Our luxury cars feature innovative designs along with legendary performance to deliver one of the top sports cars in the ...
  2014 Jaguar F-TYPE Price ... - 2013 Jaguar XF Price & Models - Models & Pricing
  23,123 people +1'd this

- Jaguar - Wikipedia, the free encyclopedia
  en.wikipedia.org/wiki/Jaguar
  The jaguar is a big cat, a feline in the Panthera genus, and is the only Panthera species found in the Americas. The jaguar is the third-largest feline after the tiger ...

- Jaguar Cars - Wikipedia, the free encyclopedia
  en.wikipedia.org/wiki/Jaguar_Cars
  Jaguar Cars since December 2012 officially incorporated as Jaguar Land Rover Ltd. is a British multinational car manufacturer headquartered in Whitley, ...

- Jaguar Dealer in Raleigh NC - New Used Jaguar Cars Durham ...
  www.leithjaguar.com
  Jaguar Dealer in Raleigh, NC. New Used & CPO Jaguar cars for sale. Jaguar Parts & service serving Raleigh Durham, Cary, Durham & Apex. Jaguar XF, XJ, XK ...

- Raleigh Used Jaguar Cars For Sale Cary Apex Durham NC
  www.leithjaguar.com/searchused.aspx
  Used & CPO Jaguar cars for sale in Raleigh NC. Used Jaguar dealer in Raleigh, Cary Apex, North Carolina. Used XJ, XK, XF & F-Type cars for sale in inventory ...

- Jaguars, Jaguar Pictures, Jaguar Facts - National Geographic
  animals.nationalgeographic.com/animals/mammals/jaguar/
  Learn all you wanted to know about Jaguars with pictures, videos, photos, facts, and news from National Geographic.

- Jaguar News - Autoblog
  www.autoblog.com > Automobiles
  Read the latest Jaguar news and reviews complete with photos, videos, and road tests of new Jaguar vehicles.

- Jacksonville Jaguars, Official Site of the Jacksonville Jaguars
  www.jaguars.com
  The official team site with scores, news items, game schedule, and roster.
Diversified results

Images might be especially helpful on SERPs with diversified results from ambiguous queries.
Prior work: images in surrogates

- Three general approaches:
  - Thumbnail of underlying page
  - Combination of visual + textual elements
  - Single images from underlying page
- Mixed results on effectiveness

Woodruff et al., 2001

Teevan, et al., 2009

Jiao et al., 2010

Aula, et al., 2010

Li, et al., 2009

Loumakis et al, 2011
Commercial efforts
Our focus

Use an image from the underlying page to **augment** the textual components of the surrogate
Individual surrogate vs. whole SERP

Two ways images could help (or hurt)

Do images help/hurt?

For individual surrogate level judgments?

For helping to make sense of the SERP as a whole?
Image quality

Not all images on a page are helpful/representative

Do images help/hurt?

For individual surrogate level judgments?

For helping to make sense of the SERP as a whole?

Does image quality matter?

Does image quality matter?

Good image = strongly related to the main content of the page

Bad image = not related to the main content of the page
Diversified vs. homogeneous results

Ambiguous queries $\rightarrow$ diversified results
Unambiguous query $\rightarrow$ more homogeneous results

Do images help/hurt?

For individual surrogate level judgments?

For helping to make sense of the SERP as a whole?

Does image quality matter?

Does image quality matter? Does diversified vs. homogeneous results matter?

Do images help more for SERPs with diversified results?
Research questions – Two user studies

1. Do images help?
2. Does image quality matter?
3. Do images help more with diversified results?

Do images help/hurt?

1. For individual surrogate level judgments?
   - Image quality
2. For helping to make sense of the SERP as a whole?
   - Image quality
3. Diversified vs. homogeneous results

Study 1: Individual surrogate judgments

Study 2: SERP-level interaction
Study overviews

Study 1 – Individual surrogates

Find information about Jake LaMotta, a boxer known as the Raging Bull.

Jake LaMotta - IMDb
http://www.imdb.com/name/nm0493796/

I think the underlying webpage would be
USEFUL  NOT USEFUL
when trying to answer this search task.

Study 2 – SERP interaction

Task: Find pricing information about the Porsche Cayenne.

cayenne

WEB RESULTS: 1 - 10 out of 50

None of these results contain the requested information.

Cayenne Models - All Porsche Vehicles - HOME - Dr. Ing. h.c. F.
Cayenne; Cayenne Diesel; Cayenne S; Cayenne S Hybrid Cayenne GTS Cayenne Turbo

Cayenne Pepper
Cayenne pepper is one of the greatest health secrets in the world. The health benefits of cayenne pepper are many including preventing heart disease, cancer, ...
www.cayennepepper.info/

Used Porsche Cayenne - Leesburg, VA - For sale on Cars.com
Search Porsche Cayenne listing by price, mileage and more. Chat with a local dealer or request a quote online.
www.cars.com/for-sale/porsche/cayenne/leesburg/

Cayenne : Herbs To Herbs
Description regarding cayenne, and how the hotness of different chilies can be measured and scored.

Cayenne pepper for improved blood flow - Healing and prevention
Information on using cayenne pepper to improve blood flow and other health benefits.
www.healingdaily.com/detoxification-dot/cayenne.htm

Cayenne pepper - Wikipedia, the free encyclopedia
The cayenne pepper also known as the Guinea spice, cow-horn pepper, alewa, bird pepper, or, especially in its powdered form, red pepper is a red, hot chili ...
www.wikipedia.org/wiki/Cayenne_pepper

Curing With Cayenne - Amazing healing power of cayenne pepper
If you master only one herb in your life, master cayenne pepper. It is more powerful than any other.
www.kholys-wellness.ca/com/cayenne.htm
Experimental Variables

• Search task
  ▪ 150 search tasks
  ▪ Ex: “Find information about the Mitsubishi Eclipse?”

• Query type
  ▪ Ambiguous → diversified results
  ▪ Unambiguous → more homogeneous results

• Image type
  ▪ Good image – strongly relates to the main focus of the page
  ▪ Bad image
  ▪ Mixed
  ▪ No images – baseline
### Task and Query Creation

<table>
<thead>
<tr>
<th>Task</th>
<th>Ambiguous query</th>
<th>Unambiguous query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find information about the Achilles tendon.</td>
<td>achilles</td>
<td>achilles tendon</td>
</tr>
<tr>
<td>Find information about the Chevrolet Corsica.</td>
<td>corsica</td>
<td>chevy corsica</td>
</tr>
<tr>
<td>Find information about the rock band Genesis.</td>
<td>genesis</td>
<td>genesis band</td>
</tr>
<tr>
<td>Find information about Swanson, a brand of frozen and canned foods.</td>
<td>swanson</td>
<td>swanson food brand</td>
</tr>
</tbody>
</table>

- Ambiguous entities identified from Wikipedia disambiguation pages
- Filtered against AOL query log
- Selected 150 popular entities and wrote task descriptions
## Task and Query Creation

<table>
<thead>
<tr>
<th>Task</th>
<th>Ambiguous query</th>
<th>Unambiguous query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find information about the Achilles tendon.</td>
<td>achilles</td>
<td>achilles tendon</td>
</tr>
<tr>
<td>Find information about the Chevrolet Corsica.</td>
<td>corsica</td>
<td>chevy corsica</td>
</tr>
<tr>
<td>Find information about the rock band Genesis.</td>
<td>genesis</td>
<td>genesis band</td>
</tr>
<tr>
<td>Find information about Swanson, a brand of frozen and canned foods.</td>
<td>swanson</td>
<td>swanson food brand</td>
</tr>
</tbody>
</table>

- Unambiguous queries collected from a crowdsourced study
- Presented task description, users issued queries
- Selected most common query out of x10 redundancy
Search Results and Images

- **Search Results**
  - Title, URL, & snippet were cached from Bing Search API
  - 150 tasks x 2 queries x 10 results = 3,000 text surrogates
  - Relevance of each underlying page was judged by trained assessors.

- **Images**
  - were cached from the underlying 3,000 web pages
  - For each page, a good and bad image was selected based on a rule-based classifier + crowdsourced judgments

---

**Used Porsche Cayenne - Leesburg, VA - For sale on Cars.com**
Search Porsche Cayenne listing by price, mileage and more. Chat with a local dealer or request a quote online.
[www.cars.com/for-sale/porsche/cayenne/leesburg/](http://www.cars.com/for-sale/porsche/cayenne/leesburg/)

**Cayenne - Herbs To Herbs**
Description regarding cayenne, and how the hotness of different chilies can be measured and scored.
[www.herb2herbs.com/herbs/herbs_cayenne.htm](http://www.herb2herbs.com/herbs/herbs_cayenne.htm)
Study 1

Find information about Jake LaMotta, a boxer known as the Raging Bull.

I think the underlying webpage would be [USEFUL] when trying to answer this search task.
Study 1 – Experimental design

- **Independent variables**
  - Task: 128 tasks (not all pages had good/bad image)
  - Query type: ambiguous, unambiguous
  - Image type: good, bad, mixed, no image

- **Fully crossed**
  - 128 x 2 x 4 x 5 (redundancy)

- **Dependent variables**
  - Accuracy = % correct
  - Average judgment duration (in milliseconds)
Study 1 – Procedure

- Crowdsourced on Amazon Mechanical Turk
  - Workers cut-off if they failed > 3 “check” tasks
  - Tasks were randomly assigned
  - Workers were randomly assigned to an image condition
  - Surrogates presented in groups of 5 to 10 for the same task

- Total judgments collected = 34,000
Study 1 – Accuracy

Results:
(based on 34,000 judgments)

• Good images led to 2.3% improvement in accuracy over text-only (p<.01).
  (Effect of image: $\chi^2(3) = 24.26$, p<.01)

• Overall, accuracy was high.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>0.864</td>
</tr>
<tr>
<td>bad</td>
<td>0.864</td>
</tr>
<tr>
<td>mixed</td>
<td>0.861</td>
</tr>
<tr>
<td>good</td>
<td>0.884</td>
</tr>
</tbody>
</table>
Study 1 – Accuracy

Some good news:

- No significant drop in accuracy for bad and mixed images.
- This means good helped (a little) and bad didn’t hurt.
Study 1 – Binned Analysis

High bin = perfect accuracy in text-only condition
Low bin = everything else

Good vs. text-only, 24% increase (p<.01)
(Effect of image: $\chi^2(3) = 119.67$, p<.01)

<table>
<thead>
<tr>
<th>Category</th>
<th>Low Bin (n=2765)</th>
<th>High Bin (n=5735)</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>0.583</td>
<td>1.000</td>
</tr>
<tr>
<td>bad</td>
<td>0.660</td>
<td>0.963</td>
</tr>
<tr>
<td>mixed</td>
<td>0.669</td>
<td>0.954</td>
</tr>
<tr>
<td>good</td>
<td>0.721</td>
<td>0.962</td>
</tr>
</tbody>
</table>
Study 1 – Binned Analysis

High bin = perfect accuracy in text-only condition
Low bin = everything else

<table>
<thead>
<tr>
<th></th>
<th>low bin (n=2765)</th>
<th>high bin (n=5735)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>text-only</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>0.583</td>
<td>0.721</td>
</tr>
<tr>
<td></td>
<td>bad</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>0.669</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>mixed</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.963</td>
</tr>
</tbody>
</table>

Good vs. text-only, 4% decrease (p < .01)
(Effect of image: $\chi^2(3) = 249.42$, $p < .01$)
Study 1 – Judgment duration

High bin = perfect accuracy in text-only condition
Low bin = everything else

Good vs. all, ~230ms increase (p<.01)
(Effect of image: F(3)=25.55, p<.01)

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>low bin</th>
<th>high bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>2313</td>
<td>2903</td>
<td>2029</td>
</tr>
<tr>
<td>bad</td>
<td>2361</td>
<td>2938</td>
<td>2084</td>
</tr>
<tr>
<td>mixed</td>
<td>2357</td>
<td>2928</td>
<td>2083</td>
</tr>
<tr>
<td>good</td>
<td>2544</td>
<td>3275</td>
<td>2193</td>
</tr>
</tbody>
</table>
Study 1 – Judgment duration

High bin = perfect accuracy in text-only condition
Low bin = everything else

Good vs. all, ~350ms increase

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>low bin</th>
<th>high bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>2313</td>
<td>2903</td>
<td>2029</td>
</tr>
<tr>
<td>bad</td>
<td>2361</td>
<td>2938</td>
<td>2084</td>
</tr>
<tr>
<td>mixed</td>
<td>2357</td>
<td>2928</td>
<td>2083</td>
</tr>
<tr>
<td>good</td>
<td>2544</td>
<td>3275</td>
<td>2193</td>
</tr>
</tbody>
</table>
Study 1 – Results Summary

Do images help/hurt?

1. Good images help, but only a small amount (2.3%)
2. Image quality (good vs. bad) matters. Bad images didn’t help at all.

Study 1: Individual surrogate judgments
Study 2

Do images help/hurt?

1. For helping to make sense of the SERP as a whole?

2. Image quality

3. Diversified vs. homogeneous results

Study 2: SERP-level interaction
Study 2

- **SERP-level judgments**
  - Initial query & SERP provided
  - Task: select a relevant result
  - Or, click the “none relevant” button
  - No query reformulation allowed
Study 2 – Experimental design

- Independent variables
  - Same as Study 1
- Fully crossed
  - 150 tasks x 2 query types x 4 image conditions x 5 redundancy
  - 6,000 total trials
- Dependent variables
  - Click-precision = # clicks on relevant results / total # clicks
  - Time-to-completion = (end time – start time)
- Crowdsourced on Amazon Mechanical Turk
  - Random assignment of tasks
  - Workers assigned a random image condition
Study 2 – Click precision

Good vs. text-only, 2% increase (not sig.)

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>ambiguous queries</th>
<th>unambiguous queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>0.865</td>
<td>0.771</td>
<td>0.931</td>
</tr>
<tr>
<td>bad</td>
<td>0.848</td>
<td>0.764</td>
<td>0.905</td>
</tr>
<tr>
<td>mixed</td>
<td>0.853</td>
<td>0.763</td>
<td>0.915</td>
</tr>
<tr>
<td>good</td>
<td>0.882</td>
<td>0.822</td>
<td>0.920</td>
</tr>
</tbody>
</table>
Study 2 – Click precision

Good vs. text-only, 7% increase
(Effect of image: $\chi^2(3) = 6.80, p = .08$)

No sig. diff

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>ambiguous queries</th>
<th>unambiguous queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>0.865</td>
<td>0.771</td>
<td>0.931</td>
</tr>
<tr>
<td>bad</td>
<td>0.848</td>
<td>0.764</td>
<td>0.905</td>
</tr>
<tr>
<td>mixed</td>
<td>0.853</td>
<td>0.763</td>
<td>0.915</td>
</tr>
<tr>
<td>good</td>
<td>0.882</td>
<td>0.822</td>
<td>0.920</td>
</tr>
</tbody>
</table>
Study 2 – Time-to-completion

Bad & mixed > text-only & good (p<.01)
(Effect of image: F(3)=9.38, p<.01)

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>ambiguous queries</th>
<th>unambiguous queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>text-only</td>
<td>29094</td>
<td>29353</td>
<td>28835</td>
</tr>
<tr>
<td>bad</td>
<td>32337</td>
<td>31407</td>
<td>33268</td>
</tr>
<tr>
<td>mixed</td>
<td>33458</td>
<td>32665</td>
<td>34251</td>
</tr>
<tr>
<td>good</td>
<td>29149</td>
<td>29111</td>
<td>29188</td>
</tr>
</tbody>
</table>
Do images help/hurt?

1. Good images may help, but only a small amount (2.0%)
2. Image quality matters.
   Bad images didn’t help click-precision at all.
   Bad images led to longer time-to-completion
3. For ambiguous queries with diversified results, good images helped more (vs. unambig/homogeneous)
Implications

- Images could help in specific cases
  - When the query is ambiguous and results are diversified
  - When the textual parts of surrogates are lacking

- Both these cases can be predicted
  - Many ambiguous queries are already known
  - Simple metrics could predict text surrogate quality

- Fast image classifiers can select a good image about 85% of the time

- Image use could be selectively, algorithmically applied
Questions

Rob Capra
@rcapra3
rcapra@unc.edu
UNC School of Information and Library Science