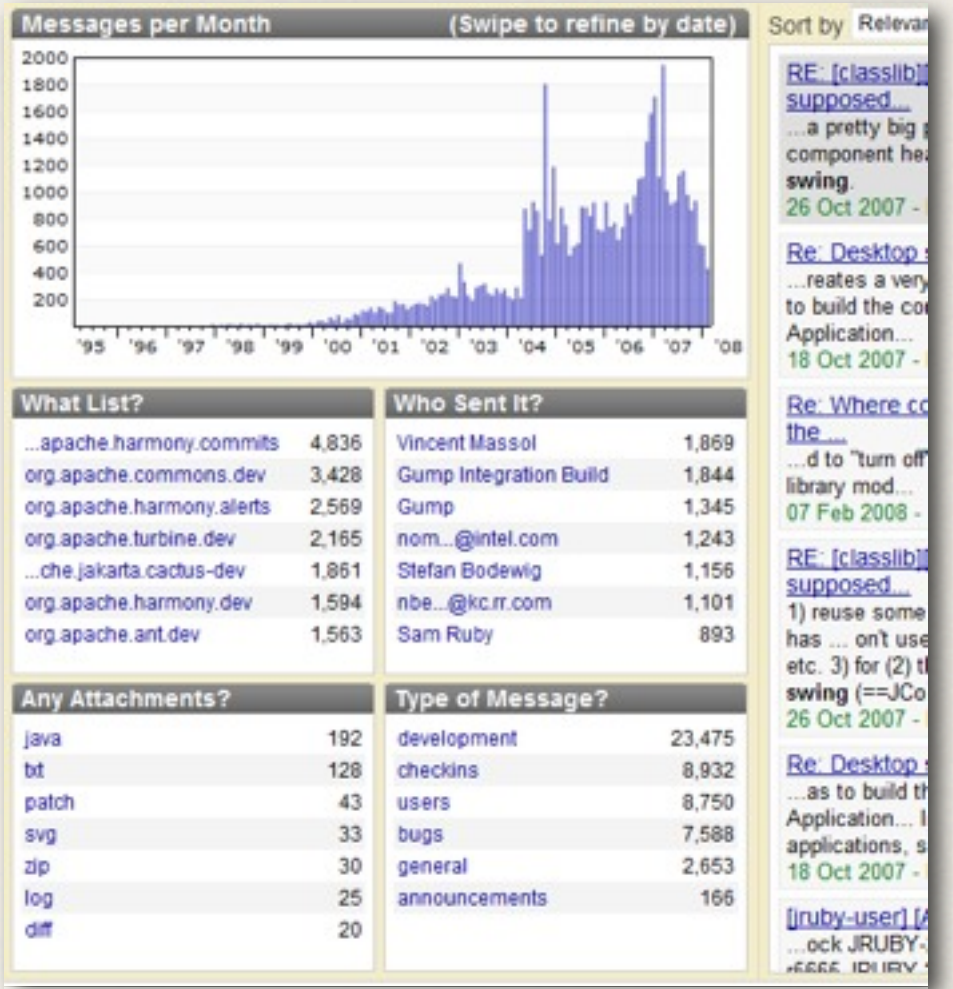
http://datamining.typepad.com/data_mining/2008/02/cifoo-markmail.html



LIS466: Ranking and IR

Week 6: Thinking about ranking functions

Today

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):
  - Query Representation

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):
  - Query Representation
  - Document Collection Representation

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):
  - Query Representation
  - Document Collection Representation
  - Matching Functions

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):
  - Query Representation
  - Document Collection Representation
  - Matching Functions
  - Ranking Functions

Today

- So far, we've reviewed some leading publications (van Rijsbergen and others)
- Divided IR into its major functions (à la Baeza-Yates):
  - Query Representation
  - Document Collection Representation
  - Matching Functions
  - Ranking Functions
- Today we look at ranking functions

Today - Ranking

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.
    - Understanding how query terms lead to a set of documents and why those documents appear in the order they do

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.
    - Understanding how query terms lead to a set of documents and why those documents appear in the order they do
- From here, let's touch every so briefly on the human's understanding of ranking

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.
    - Understanding how query terms lead to a set of documents and why those documents appear in the order they do
- From here, let's touch every so briefly on the human's understanding of ranking
  - Interfaces and websites

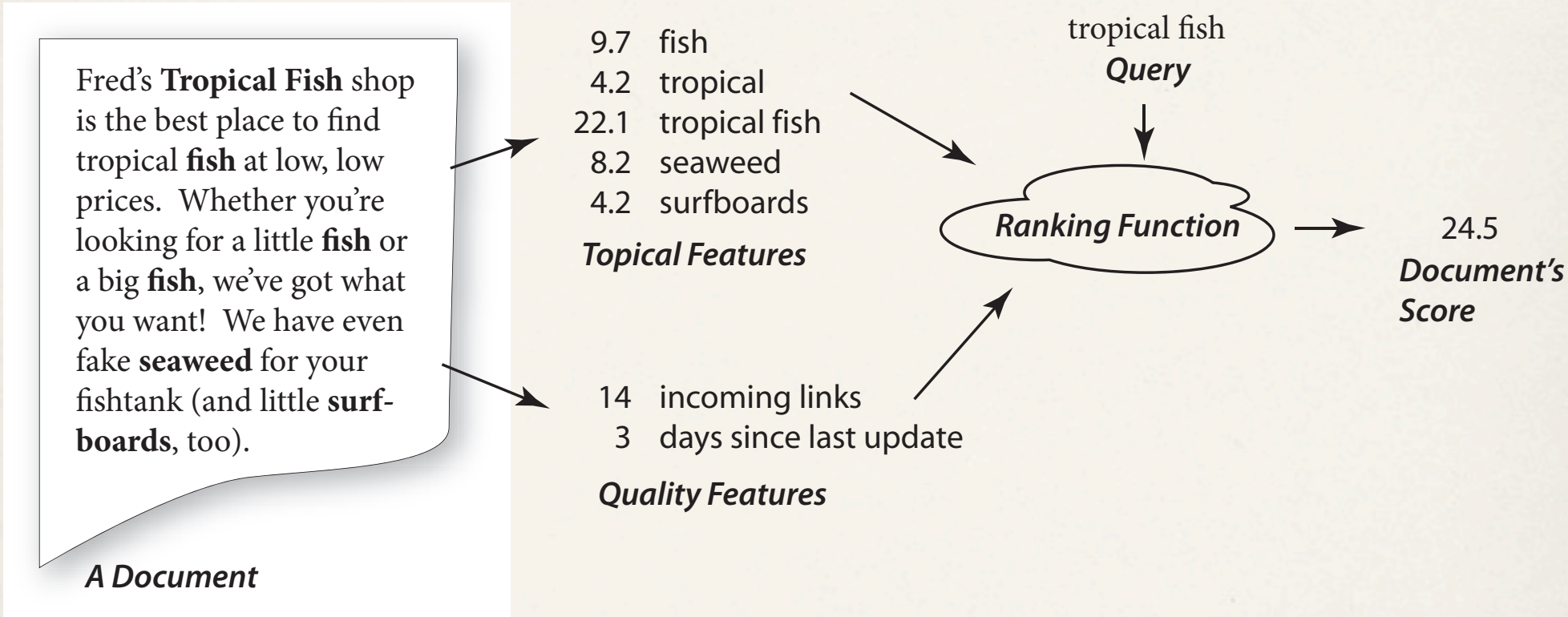
Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.
    - Understanding how query terms lead to a set of documents and why those documents appear in the order they do
- From here, let's touch every so briefly on the human's understanding of ranking
  - Interfaces and websites
  - A necessary touch on the topic of data mining

Today - Ranking

- Let's consider Ranking as having multiple dimensions in IR
  - From the machine's p.o.v.
    - Computational efficiency
  - From the human user's p.o.v.
    - Understanding how query terms lead to a set of documents and why those documents appear in the order they do
- From here, let's touch every so briefly on the human's understanding of ranking
  - Interfaces and websites
  - A necessary touch on the topic of data mining
- By the end of today, you should have reflected on issues of ranking and considered how the question of ranking remains open.

Abstract model of ranking



In this example, we see how features from the document are extracted, weighted, and collectively contribute to a doc's score

file: abstractRanking1

Remember...

Remember...

- Documents are parsed and the text transformed into *index terms* or *document features*.

Remember...

- Documents are parsed and the text transformed into *index terms* or *document features*.
- Recall, too, that we can add other forms of evidence or features, such as currency, to weighting and classification of documents. In the previous slide, there's a reference to the document's last update being 3 days ago.

Remember...

- Documents are parsed and the text transformed into *index terms* or *document features*.
- Recall, too, that we can add other forms of evidence or features, such as currency, to weighting and classification of documents. In the previous slide, there's a reference to the document's last update being 3 days ago.
  - If currency were important in your IR system then you'd rank something lower the more days away from the query date.

Remember...

- Documents are parsed and the text transformed into *index terms* or *document features*.
- Recall, too, that we can add other forms of evidence or features, such as currency, to weighting and classification of documents. In the previous slide, there's a reference to the document's last update being 3 days ago.
  - If currency were important in your IR system then you'd rank something lower the more days away from the query date.
- In the previous example, the "ranking function" was a featureless cloud - don't worry - the point is that some score is generated. The closer numerically that document score is to the query score, the argument goes, the two are closely related.

Remember...

- Documents are parsed and the text transformed into *index terms* or *document features*.
- Recall, too, that we can add other forms of evidence or features, such as currency, to weighting and classification of documents. In the previous slide, there's a reference to the document's last update being 3 days ago.
  - If currency were important in your IR system then you'd rank something lower the more days away from the query date.
- In the previous example, the "ranking function" was a featureless cloud - don't worry - the point is that some score is generated. The closer numerically that document score is to the query score, the argument goes, the two are closely related.
  - Of course, this begs the question of whether the designers' choice of features *and* the semantic power of terms really are useful.

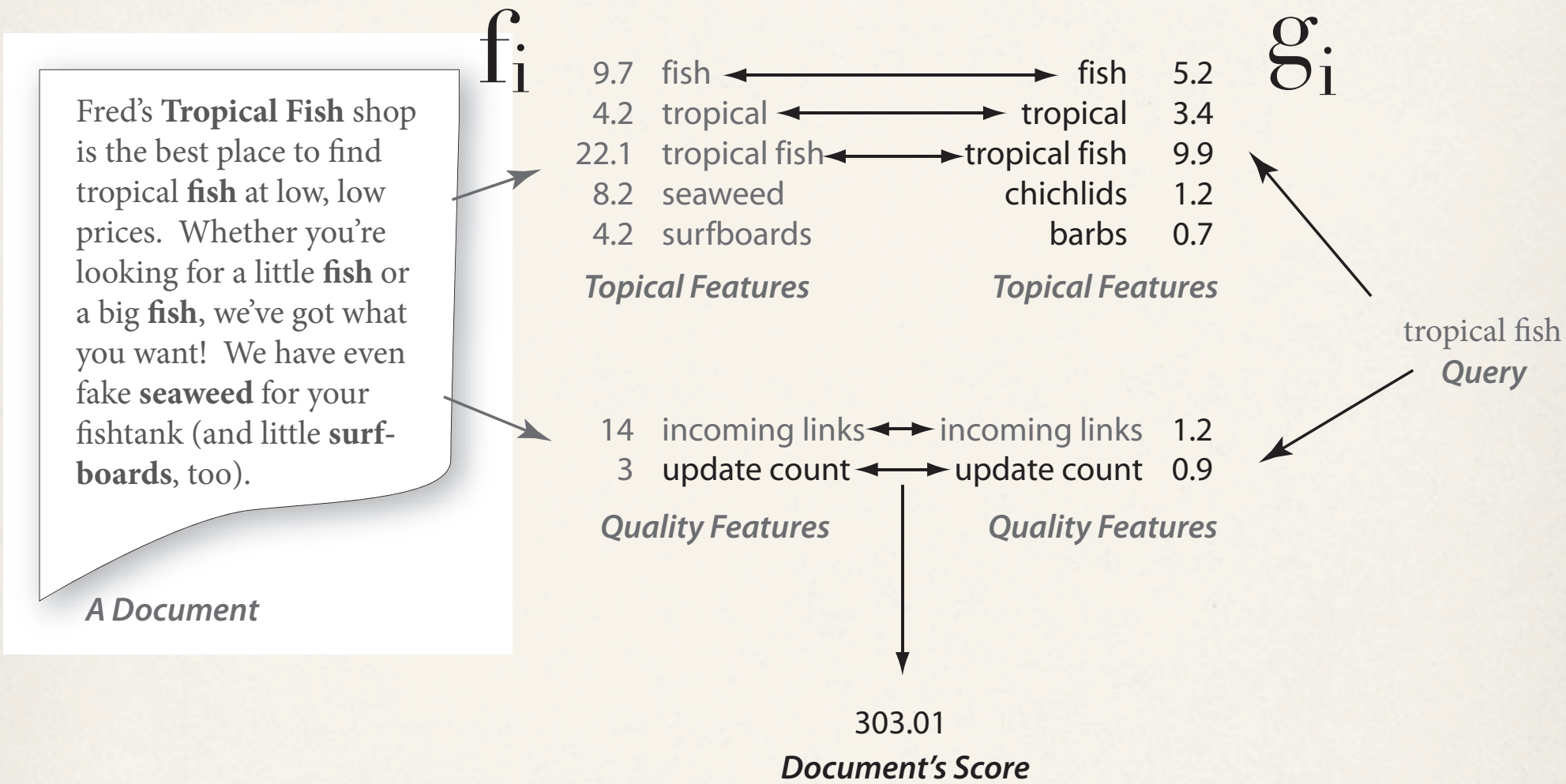
A more concrete example

- Typically a ranking function's ability to match with discrimination the query to the document representation is the measure also of its effectiveness. Different search engines are often compared using the same test collection to see which algorithm is more efficient.
- Here's a typical model of a ranking function R:

$$R(Q,D)=\sum_i g_i(Q)f_i(D)$$

- Here, f_i is some feature function that extracts a number from the document text. g_i is a similar feature function that extracts a value from a query. These two functions form a pair of feature functions. Each pair is multiplied together, and the results from all pairs are added to create a final document.

Model using query & document features



Both query & document have feature functions

file: abstractRanking2

Notes about the concrete model

- Notice that $f_i(D)$ can be replaced with the terms to help clarify things:
- $f_i(D)$ becomes $f_{tropical}(Fred's Ad)$ or $f_{fish}(D)$
- The Query aspect represented by $g_{tropical}(Q)$ evaluates a large value because "tropical" is in the query. However, $g_{barbs}(Q)$ also has a small non-zero value because it is *related* to other terms in the query.
- The values are all summarized to form a score for the document.
- BTW: If the query were "today's weather in Boston" we would prefer documents that are updated frequently, since a document about the weather changes a lot; compare this to "full text of Moby Dick" probably doesn't get updated so update feature would make a low value for the novel but higher value for weather-related documents.

Inverted files (aka indexes aka indices)

- As you know, an index is *n* number of characters of the terms along with a document pointer all combined with some technique for fast examination of the index.
- For instance, let's say we have the terms "cat fish documentary chowder" all in Document #1 and our index is a bi-gram:

```
ca  1
fi  1
do  1
ch  1
```

Example (Three Sentences S_1, S_2, S_3)

S_1 Tropical fish include bright fish found in tropical environments around the world, including both freshwater and salt water species.
S2 Fishkeepers often use the term tropical fish to refer only those requiring fresh water; with saltwater tropical fish referred to as marine fish.
S3 Tropical fish are popular aquarium fish, due to their often bright coloration.

and	1
aquarium	3
are	3
around	1
as	2
both	1
bright	1,3
coloration	3
due	3
environments	1
fish	1,2,3,4
fishkeepers	2

On the left is an inverted index for documents (sentences). In this case the "document" is a sentence long.

What to do with inverted indices?

What to do with inverted indices?

- The index provides the very least a binary match: the term is there or it isn't.

What to do with inverted indices?

- The index provides the very least a binary match: the term is there or it isn't.
- But usually people consider the intersection of term: For example "coloration" appears in only S_3; "bright" appears in S_1 and S_3.

What to do with inverted indices?

- The index provides the very least a binary match: the term is there or it isn't.
- But usually people consider the intersection of term: For example "coloration" appears in only S_3; "bright" appears in S_1 and S_3.
- To rank documents rather than just noting their presence (binary match), we need another technique.

What to do with inverted indices?

- The index provides the very least a binary match: the term is there or it isn't.
- But usually people consider the intersection of term: For example "coloration" appears in only S_3; "bright" appears in S_1 and S_3.
- To rank documents rather than just noting their presence (binary match), we need another technique.
  - We could do anything, but let's try recording the number of times a term appears in a document, so we must record both document # and term frequency.

What to do with inverted indices?

- The index provides the very least a binary match: the term is there or it isn't.
- But usually people consider the intersection of term: For example "coloration" appears in only S_3; "bright" appears in S_1 and S_3.
- To rank documents rather than just noting their presence (binary match), we need another technique.
  - We could do anything, but let's try recording the number of times a term appears in a document, so we must record both document # and term frequency.
  - We could also record the *position* of the term in the document. The position could be some known part of the document (such as the title, abstract) *or* it could be the position of term follows the position of another term, say "fish" following a section about "seawater".

Example of document & position table

and	1:1
aquarium	3:1
are	3:1
around	1:1
as	2:1
both	1:1
bright	1:1,3:1
coloration	3:1
due	3:1
environments	1:1
fish	1:1,2:1,3:1,4:1
fishkeepers	2:1

Here, the document number is first followed by a colon : and then the term count.

document:count

Example with word positions

and	1,15
aquarium	3,5
are	3,3; 4,14
around	1,9
as	2,21
both	1,13
bright	1,4; 3,11
coloration	3,12
due	3,7
environments	1,8
fish	1,2; 1,4; 2,7; 2,18; 2,23; 3,2; 3,6
fishkeepers	2:1

Here the term counts are limited to one posting but we record the document number first, followed by a word position.

term document:position

Notice how much more data we have to work with about the terms!

A note about positions...

A note about positions...

- Depending on implementation, we might have a system that looks for related terms (find "marine fish") that looks for concepts related to marine (e.g., "seawater") and then looks for terms related to, and the query itself, ("fish"). The idea is that if you find fish *following* a section about *seawater* (a related term), that document or that part of a document is *more relevant to the query* than a document where "fish", "marine", and related terms appear anywhere.

A note about positions...

- Depending on implementation, we might have a system that looks for related terms (find "marine fish") that looks for concepts related to marine (e.g., "seawater") and then looks for terms related to, and the query itself, ("fish"). The idea is that if you find fish *following* a section about *seawater* (a related term), that document or that part of a document is *more relevant to the query* than a document where "fish", "marine", and related terms appear anywhere.
- Rank document D higher when term t_1 appears in the context of related term r_1, r_2, … r_n. Sum them to create a rank for that document. Compare that rank value to the query's own rank.

A note about positions...

- Depending on implementation, we might have a system that looks for related terms (find "marine fish") that looks for concepts related to marine (e.g., "seawater") and then looks for terms related to, and the query itself, ("fish"). The idea is that if you find fish *following* a section about *seawater* (a related term), that document or that part of a document is *more relevant to the query* than a document where "fish", "marine", and related terms appear anywhere.
- Rank document D higher when term t_1 appears in the context of related term r_1, r_2, … r_n. Sum them to create a rank for that document. Compare that rank value to the query's own rank.
- Ranking a query sounds odd, doesn't it? But it is just a numeric value created by some algorithm extracting terms and weighting them, too, just as we do for documents. Then $Dt \approx Qt$ [document terms values are close to the query terms value.]

Aligning the phrase

tropical	1,1			2,6	2,17		3,1		
fish	1,2	1,4	1,7	2,7	2,18	2,33	3,2	3,6	and so on...

Aligning the postings lists for "tropic" and "fish" to find the phrase "tropical fish"

Applying this ...

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

Applying this ...

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- If we can record document number, position number, and term … what else can we do?

Applying this ...

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- If we can record document number, position number, and term … what else can we do?
- In this example, searching for "aaron burr" might be treated as "aaron" + "burr". But we want the concept "aaron burr" - so want the terms next to each other ("aaron burr") to be treated differently. This is called an *extent*.

Applying this ...

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- If we can record document number, position number, and term … what else can we do?
- In this example, searching for "aaron burr" might be treated as "aaron" + "burr". But we want the concept "aaron burr" - so want the terms next to each other ("aaron burr") to be treated differently. This is called an *extent*.
- An *extent* or *extent list* is a contiguous region of a document.

Applying this ...

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- If we can record document number, position number, and term … what else can we do?
- In this example, searching for "aaron burr" might be treated as "aaron" + "burr". But we want the concept "aaron burr" - so want the terms next to each other ("aaron burr") to be treated differently. This is called an *extent*.
- An *extent* or *extent list* is a contiguous region of a document.
- So imagine would could parse the document, skipping the XML tags, and record something like author: (1,2), (3,4) that would mean "the author concept is stored in the first word (1) and goes to word #2 (1 = "Rhéaume", 2 = "Barrette"); the (3,4) means start at word 3 (Aaron) and continue to word 4 (Burr).

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- We would have to record the XML tag (such as author, title) and the content.

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- We would have to record the XML tag (such as author, title) and the content.
- Usually IR turns now to compression algorithms, which emphasize the computer's efficiency.

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- We would have to record the XML tag (such as author, title) and the content.
- Usually IR turns now to compression algorithms, which emphasize the computer's efficiency.
- For many demo projects, we use a relational database to store our data, be they in fields whose name is the same as the XML tag and data are the counts (as the previous slides suggest) or the actual term and the calculations are performed after a query has been submitted.

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- We would have to record the XML tag (such as author, title) and the content.
- Usually IR turns now to compression algorithms, which emphasize the computer's efficiency.
- For many demo projects, we use a relational database to store our data, be they in fields whose name is the same as the XML tag and data are the counts (as the previous slides suggest) or the actual term and the calculations are performed after a query has been submitted.
  - There are pros and cons to this approach. Calculating weights in response to a query is computationally intensive and probably requires calculating relevancy ranks only from documents that have at least one related term.

Applying this to XML

<author>Rhéaume Barrette</author>
<author>Aaron Burr</author>
<title>Chez les malécites</title>

- We would have to record the XML tag (such as author, title) and the content.
- Usually IR turns now to compression algorithms, which emphasize the computer's efficiency.
- For many demo projects, we use a relational database to store our data, be they in fields whose name is the same as the XML tag and data are the counts (as the previous slides suggest) or the actual term and the calculations are performed after a query has been submitted.
  - There are pros and cons to this approach. Calculating weights in response to a query is computationally intensive and probably requires calculating relevancy ranks only from documents that have at least one related term.
  - There are models that work on this principle, so let's do it, too!

The User and Ranking

The User and Ranking

- Do end-users actually think about “relevancy ranking”?

The User and Ranking

- Do end-users actually think about “relevancy ranking”?
  - Probably not - the popular idea is that search engines have hits or they’re lousy search engines.

The User and Ranking

- Do end-users actually think about "relevancy ranking"?
  - Probably not - the popular idea is that search engines have hits or they're lousy search engines.
- Are there "cognitive" and/or "affective" and even social issues?

The User and Ranking

- Do end-users actually think about "relevancy ranking"?
  - Probably not - the popular idea is that search engines have hits or they're lousy search engines.
- Are there "cognitive" and/or "affective" and even social issues?
  - What do you think that people recognize in the retrieval set that attract them? [Really, think about this for a moment; perhaps do some searches on different systems. What do *you* identify?]

The User and Ranking

- Do end-users actually think about "relevancy ranking"?
  - Probably not - the popular idea is that search engines have hits or they're lousy search engines.
- Are there "cognitive" and/or "affective" and even social issues?
  - What do you think that people recognize in the retrieval set that attract them? [Really, think about this for a moment; perhaps do some searches on different systems. What do *you* identify?]
  - Are there visual or text features that draw the user in? Flashing ads, pop-under pages, ads and links that have nothing to do with the query are visual noise that are known to reduce the user's sense of relevancy.

The User and Ranking

- Do end-users actually think about "relevancy ranking"?
  - Probably not - the popular idea is that search engines have hits or they're lousy search engines.
- Are there "cognitive" and/or "affective" and even social issues?
  - What do you think that people recognize in the retrieval set that attract them? [Really, think about this for a moment; perhaps do some searches on different systems. What do *you* identify?]
  - Are there visual or text features that draw the user in? Flashing ads, pop-under pages, ads and links that have nothing to do with the query are visual noise that are known to reduce the user's sense of relevancy.
  - Social issues? Peer-pressure to search certain sites other than others.

The User and Ranking

- Do end-users actually think about "relevancy ranking"?
  - Probably not - the popular idea is that search engines have hits or they're lousy search engines.
- Are there "cognitive" and/or "affective" and even social issues?
  - What do you think that people recognize in the retrieval set that attract them? [Really, think about this for a moment; perhaps do some searches on different systems. What do *you* identify?]
  - Are there visual or text features that draw the user in? Flashing ads, pop-under pages, ads and links that have nothing to do with the query are visual noise that are known to reduce the user's sense of relevancy.
  - Social issues? Peer-pressure to search certain sites other than others.
  - Technical? The way Microsoft "offers" to save passwords, or perform functions for users, who just get beaten down by confusing settings and preference files and yield.

In-class Activity

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.
- What kind of "related searches" were offered?

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.
- What kind of "related searches" were offered?
  - How do you think those terms were selected?

In-class Activity

- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.
- What kind of "related searches" were offered?
  - How do you think those terms were selected?
- How was your query treated? Were you offered "corrections"?

In-class Activity

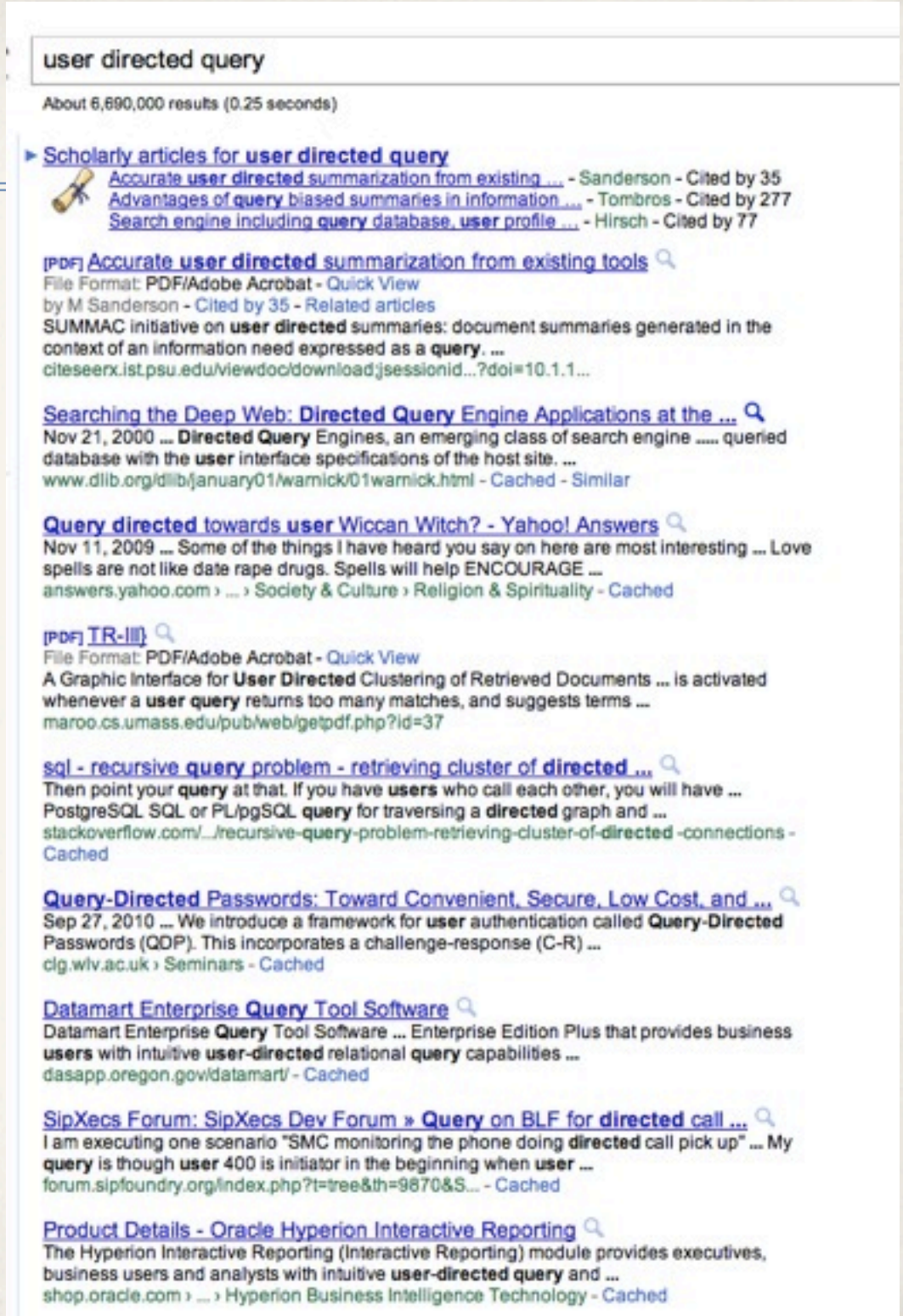
- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.
- What kind of "related searches" were offered?
  - How do you think those terms were selected?
- How was your query treated? Were you offered "corrections"?
- For discussion's sake, compare your search output to the ones that appear on the next screens. Do the same documents appear? In the same order?

In-class Activity

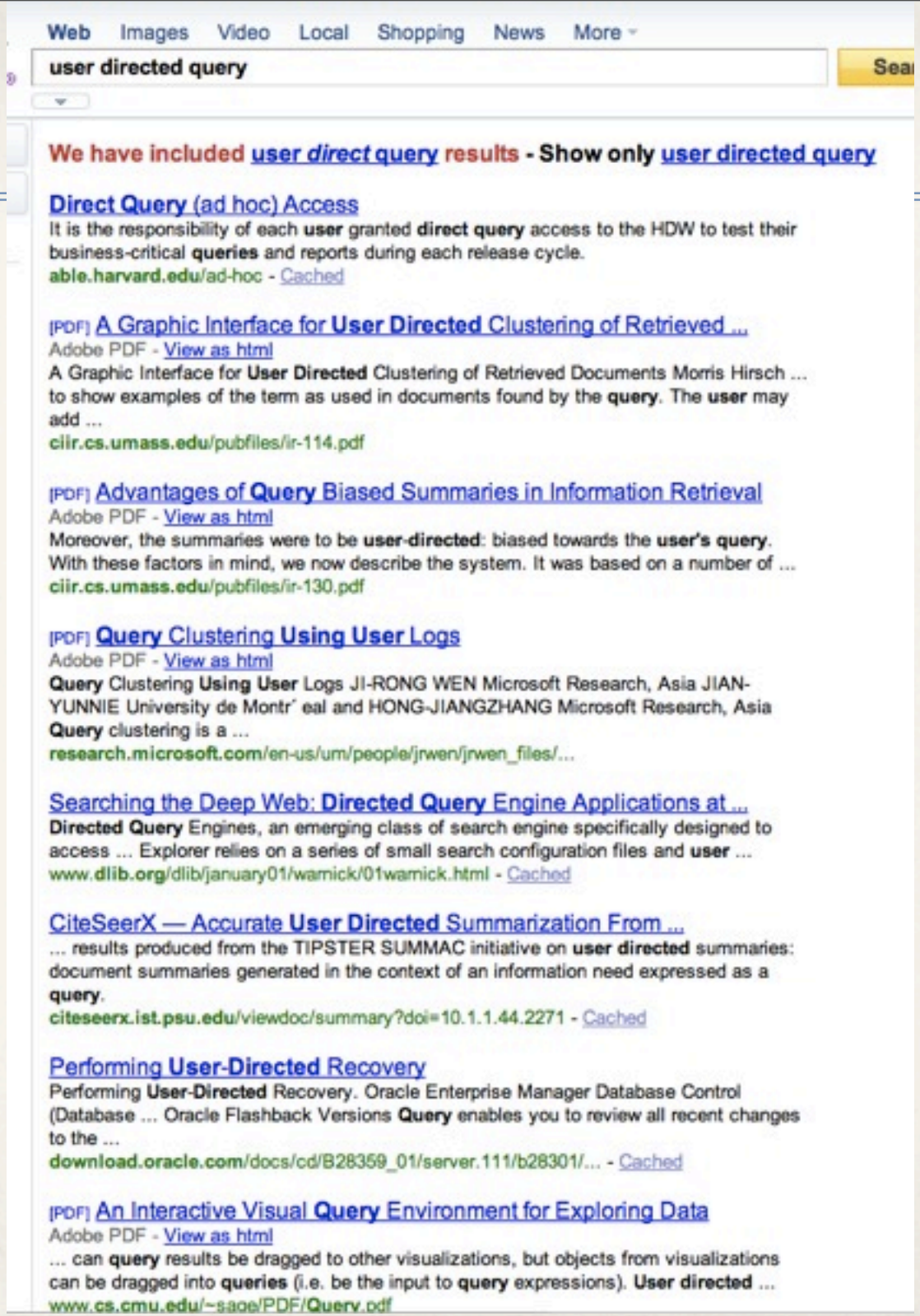
- *Before continuing* search for the phrase "user directed query" (no quotes) in each of these web-based search engines [**www.google.com**, **www.yahoo.com**, **www.bing.com**]. Save your output to compare to the following slides.
- On the next 3 slides I have the output from the search I conducted on Feb 28, 2011, 3:46 pm, from Cambridge, MA, using Comcast as the ISP.
- The Bing search shows that I'm located in "Springfield, Massachusetts". (I'm certainly *not* there!). Google and Yahoo pinpoint me to Cambridge.
- What kind of "related searches" were offered?
  - How do you think those terms were selected?
- How was your query treated? Were you offered "corrections"?
- For discussion's sake, compare your search output to the ones that appear on the next screens. Do the same documents appear? In the same order?
- Try searching with quote marks; or try forcing boolean "AND" using the +

Output from Google

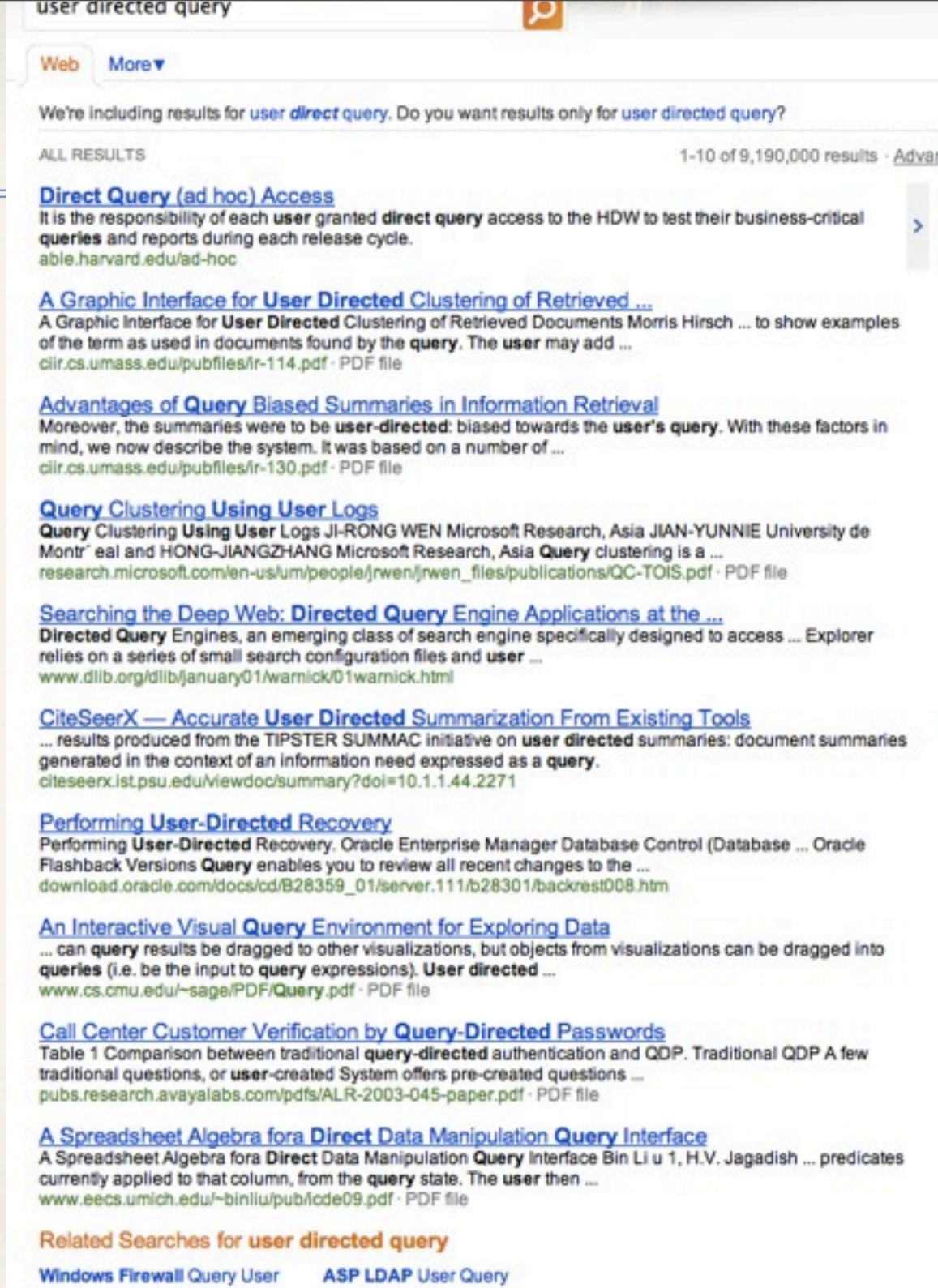
A comparison chart appears after the screen shots.



Yahoo


Web Images Video Local Shopping News More
user directed query
We have included user direct query results - Show only user directed query
Direct Query (ad hoc) Access
It is the responsibility of each user granted direct query access to the HDW to test their business-critical queries and reports during each release cycle.
able.harvard.edu/ad-hoc - Cached
[PDF] A Graphic Interface for User Directed Clustering of Retrieved ...
Adobe PDF - View as html
A Graphic Interface for User Directed Clustering of Retrieved Documents Morris Hirsch ... to show examples of the term as used in documents found by the query. The user may add ...
ciir.cs.umass.edu/pubfiles/ir-114.pdf
[PDF] Advantages of Query Biased Summaries in Information Retrieval
Adobe PDF - View as html
Moreover, the summaries were to be user-directed: biased towards the user's query. With these factors in mind, we now describe the system. It was based on a number of ...
ciir.cs.umass.edu/pubfiles/ir-130.pdf
[PDF] Query Clustering Using User Logs
Adobe PDF - View as html
Query Clustering Using User Logs JI-RONG WEN Microsoft Research, Asia JIAN-YUNNIE University de Montr´ eal and HONG-JIANGZHANG Microsoft Research, Asia Query clustering is a ...
research.microsoft.com/en-us/um/people/jrwen/jrwen_files/...
Searching the Deep Web: Directed Query Engine Applications at ...
Directed Query Engines, an emerging class of search engine specifically designed to access ... Explorer relies on a series of small search configuration files and user ...
www.dlib.org/dlib/january01/warnick/01warnick.html - Cached
CiteSeerX — Accurate User Directed Summarization From ...
... results produced from the TIPSTER SUMMAC initiative on user directed summaries: document summaries generated in the context of an information need expressed as a query.
citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.2271 - Cached
Performing User-Directed Recovery
Performing User-Directed Recovery. Oracle Enterprise Manager Database Control (Database ... Oracle Flashback Versions Query enables you to review all recent changes to the ...
download.oracle.com/docs/cd/B28359_01/server.111/b28301/... - Cached
[PDF] An Interactive Visual Query Environment for Exploring Data
Adobe PDF - View as html
... can query results be dragged to other visualizations, but objects from visualizations can be dragged into queries (i.e. be the input to query expressions). User directed ...
www.cs.cmu.edu/~sage/PDF/Query.pdf

Bing


user directed query
Web
More▾
We're including results for user direct query. Do you want results only for user directed query?
ALL RESULTS
1-10 of 9,190,000 results · Advar
Direct Query (ad hoc) Access
It is the responsibility of each user granted direct query access to the HDW to test their business-critical queries and reports during each release cycle.
able.harvard.edu/ad-hoc
A Graphic Interface for User Directed Clustering of Retrieved ...
A Graphic Interface for User Directed Clustering of Retrieved Documents Morris Hirsch ... to show examples of the term as used in documents found by the query. The user may add ...
ciir.cs.umass.edu/pubfiles/ir-114.pdf · PDF file
Advantages of Query Biased Summaries in Information Retrieval
Moreover, the summaries were to be user-directed: biased towards the user's query. With these factors in mind, we now describe the system. It was based on a number of ...
ciir.cs.umass.edu/pubfiles/ir-130.pdf · PDF file
Query Clustering Using User Logs
Query Clustering Using User Logs JI-RONG WEN Microsoft Research, Asia JIAN-YUNNIE University de Montr´ eal and HONG-JIANGZHANG Microsoft Research, Asia Query clustering is a ...
research.microsoft.com/en-us/um/people/jrwen/jrwen_files/publications/QC-TOIS.pdf · PDF file
Searching the Deep Web: Directed Query Engine Applications at the ...
Directed Query Engines, an emerging class of search engine specifically designed to access ... Explorer relies on a series of small search configuration files and user ...
www.dlib.org/dlib/january01/warnick/01warnick.html
CiteSeerX — Accurate User Directed Summarization From Existing Tools
... results produced from the TIPSTER SUMMAC initiative on user directed summaries: document summaries generated in the context of an information need expressed as a query.
citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.2271
Performing User-Directed Recovery
Performing User-Directed Recovery. Oracle Enterprise Manager Database Control (Database ... Oracle Flashback Versions Query enables you to review all recent changes to the ...
download.oracle.com/docs/cd/B28359_01/server.111/b28301/backrest008.htm
An Interactive Visual Query Environment for Exploring Data
... can query results be dragged to other visualizations, but objects from visualizations can be dragged into queries (i.e. be the input to query expressions). User directed ...
www.cs.cmu.edu/~sage/PDF/Query.pdf · PDF file
Call Center Customer Verification by Query-Directed Passwords
Table 1 Comparison between traditional query-directed authentication and QDP. Traditional QDP A few traditional questions, or user-created System offers pre-created questions ...
pubs.research.avayalabs.com/pdfs/ALR-2003-045-paper.pdf · PDF file
A Spreadsheet Algebra fora Direct Data Manipulation Query Interface
A Spreadsheet Algebra fora Direct Data Manipulation Query Interface Bin Li u 1, H.V. Jagadish ... predicates currently applied to that column, from the query state. The user then ...
www.eecs.umich.edu/~binliu/pub/icde09.pdf · PDF file
Related Searches for user directed query
Windows Firewall Query User
ASP LDAP User Query

Google	Yahoo	Bing
Scholarly articles Sanderson "Accurate user directed . . ." Tombros "Advantages of query based su" Hirsch "Search engine including query. . ."		
[PDF] Accurate user directed summarization	Direct Query (ad hoc) Access	Direct Query (ad hoc) Access
Searching the Dept Web: directed query eng	[PDF] Graphic Interface for User Directed . . .	A Graphic Interface for User Directed
Query directed towards user Wiccan Witch? -	[PDF] Advantages of Query Based Summaries	Advantages of Query Based Summaries
[PDF] TR-III}	[PDF] Query Clustering Using User Logs	Query Clustering User User Logs
sql-recursive query problem - retrieving clus	Search the Deep Web	Searching the Deep Web
Query-Directed Passwords	CiteSeerX - Accurate User Directed Summarization	CiteSeerX-Accurate User
Datamart Enterprise Query Tool Software	Performing User-Directed Recovery	Performing User-Directed
SipXecs Forum: SipXecs Dev Forum	[PDF] An Interactive Visual Query Environment	An Interactive Visual Query Env
Product Details - Oracle Hyperion		Call Center Customer Verification
		A Spreadsheet Algebra for a Direct Data

After Google's Scholarly Articles links, notice the ranking...

In-class Activity

In-class Activity

- How would you theorize the differences in these search engines' behavior?

In-class Activity

- How would you theorize the differences in these search engines' behavior?

- If you were asked "which search engine is better?", how would you reply? What types of justification would you give?

In-class Activity

- How would you theorize the differences in these search engines' behavior?
- If you were asked "which search engine is better?", how would you reply? What types of justification would you give?
- How would you manipulate the search engine behavior?

In-class Activity

- How would you theorize the differences in these search engines' behavior?

- If you were asked "which search engine is better?", how would you reply? What types of justification would you give?

- How would you manipulate the search engine behavior?

- As a way of applying relevancy ranking, think about our project. We discussed having "boolean searches" [known-entity searches by drop down box and possibly icons that map to controlled vocabularies] and searches that were more "fuzzy" [queries that act more like full-text queries, but use parsed MARC records with weights assigned to different elements (we did this last week in class)].

In-class Activity

- How would you theorize the differences in these search engines' behavior?
- If you were asked "which search engine is better?", how would you reply? What types of justification would you give?
- How would you manipulate the search engine behavior?
- As a way of applying relevancy ranking, think about our project. We discussed having "boolean searches" [known-entity searches by drop down box and possibly icons that map to controlled vocabularies] and searches that were more "fuzzy" [queries that act more like full-text queries, but use parsed MARC records with weights assigned to different elements (we did this last week in class)].
- Now we can consider how we'd associate terms. Sketch out some ideas for our next session about relevancy ranking.

Special note: a real parser.

- Parsing MARC records:
  - here's a link to the source code of a parser I wrote that looks at the tags and processes them into one of 15 elements.
  - This code could be (and shall be) updated to match our class's concept of weighting based on tags and full-texts.
  - Are you thinking of interfaces for displaying the data?
  - The code is at http://web.simmons.edu/~benoit/LIS466/tagtest4.java

Ranking & Interactivity

Ranking & Interactivity

- We mentioned, too, the idea of adjusting weights in real-time.

Ranking & Interactivity

- We mentioned, too, the idea of adjusting weights in real-time.
- Your assignment (not a graded activity, but definitely something useful for us) is to

Ranking & Interactivity

- We mentioned, too, the idea of adjusting weights in real-time.
- Your assignment (not a graded activity, but definitely something useful for us) is to
  - search the research literature for research about our kind of project

Ranking & Interactivity

- We mentioned, too, the idea of adjusting weights in real-time.
- Your assignment (not a graded activity, but definitely something useful for us) is to
  - search the research literature for research about our kind of project
  - What articles do you think are useful in explaining and/or evaluating icon or image-driven user-guided or user-directed interfaces for IR?

Ranking & Interactivity

- We mentioned, too, the idea of adjusting weights in real-time.
- Your assignment (not a graded activity, but definitely something useful for us) is to
  - search the research literature for research about our kind of project
  - What articles do you think are useful in explaining and/or evaluating icon or image-driven user-guided or user-directed interfaces for IR?
  - If possible, find some links or articles and send 'em to all the class (for example, save the .pdf on your website and send us the link)

Finally ...

Finally ...

- Think about how you'd relevancy rank documents where the query and the document collection is more than one language.

Finally ...

- Think about how you'd relevancy rank documents where the query and the document collection is more than one language.
- How would including icons or images in the retrieval set affect the end-user's relevancy judgments?

Finally ...

- Think about how you'd relevancy rank documents where the query and the document collection is more than one language.
- How would including icons or images in the retrieval set affect the end-user's relevancy judgments?
  - These and the earlier questions are not wastes of time - they're useful activities and I hope you'll contribute your ideas in class over the next four weeks (XML, CLIR, Project Discussion, other projects).

Finally ...

- Think about how you'd relevancy rank documents where the query and the document collection is more than one language.
- How would including icons or images in the retrieval set affect the end-user's relevancy judgments?
  - These and the earlier questions are not wastes of time - they're useful activities and I hope you'll contribute your ideas in class over the next four weeks (XML, CLIR, Project Discussion, other projects).
- *As always, I hope you're enjoying and learning from our examinations of IR. If you have questions, ideas, concerns, exciting discoveries, send 'em along! Bye - GB*