Findability and Usability
Lessons Learnt from Text Analytics

Anna Divoli
@annadivoli
Pingar Research

CS4HS 2013 @ Unitec
Google, September 1998

Twitter, July 2006

http://www.laurencecaro.co.uk/other/what-the-worlds-biggest-websites-looked-like-at-launch/
Welcome to Thefacebook!

Thefacebook is an online directory that connects people through social networks at colleges.

We have opened up Thefacebook for popular consumption at Harvard University.

You can use Thefacebook to:
- Search for people at your school
- Find out who are in your classes
- Look up your friends' friends
- See a visualization of your social network

To get started, click below to register. If you have already registered, you can log in.

Yahoo!, March 1995

Facebook, February 2004
A Guide to Facebook’s Home Page

A simplified design provides easy access to your entire Facebook experience.

- Links to important account settings, logos, the Help Center and more.
- Reminders for birthdays and events.
- "Facebook Out" button.
Often there is resistance!
Users adapt & evolve as well!

15 YEARS AGO

Ding!!
You've got mail!!

NOW

436 unread emails
Outline

• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell
• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features
• Take away message
Outline

- Familiar Interfaces
- User Centered Design & Usability
  - Who am I & who is Pingar
  - Text analytics in a nutshell
  - Our usability studies & findings
    - 1. Search interface features: Biomedical Scientists
    - 2. Study existing popular systems
    - 3. Mock ups of specific features
- Take away message
Usability testing is important! Needs to be done right!

Boss, I finished that algorithm you asked for!

Did you do the usability evaluation?

That's right. I forgot. I'll get right to it.

Good.

Was the algorithm easy to use?

No.

Excellent!
User Centered Design

- Design
- Analysis
- Prototype
- User Feedback
User Centered Design – What should companies focus on!

FOCUS

Users

Competitors

http://headrush.typepad.com/photos/uncategorized/focusonusersnotcompetitors.jpg
http://headrush.typepad.com/creating_passionate_users/2006/07/
Usability testing tools

• Eye tracking
• Mouse clicks
• Tasks (goal & time)
• Questionnaires
• Observation
• Talk aloud
• Brain scanners

Measurements & Preferences on:

• Aesthetics
• Craftsmanship
• Technical performance
• Ergonomics
• Usability
Eye tracking
• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell
• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features
• Take away message
Who am I?

Anna Divoli
CS4HS 2013 @ Unitec

UK
BSc (1999) Biomedical Sciences
MSc (2001) Biosystems & Informatics
PhD (2006) Biomedical Text Mining:
- Sentence extraction
- Automatic protein family database annotation
- Document clustering and visualization

USA (2006-2011)
Postdoctoral Academic Research:
- Text analytics
- Information retrieval
- User search interfaces
- Usability research
- Knowledge acquisition
- Expert opinions analysis

NZ (2011- )
Pingar Research:
- Entity Extraction
- Automatic Taxonomy Generation
- Taxonomy Editor (UI)
- Friendly UIs
- Usability & UX
Who is Pingar?

A small global company that uses state of the art text analytics to offer solutions!

We analyse unstructured text!

**Entity Extraction**
People, Organizations, Locations
Dates, Money amounts, Email addresses...
+ Custom Entities

**Keyword Extraction**

**Taxonomy Term Extraction**
Domain specific terminology (can plug in any taxonomy/ontology)

**Summarization**
Select length and focus

**Redaction & Sanitization**
Ensure anonymization of sensitive information.

In English, German, French, Spanish, Dutch, Chinese, Japanese
• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell
• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features
• Take away message
A brief Text Analytics Overview – just Entity Recognition basics...
Find and classify names...

S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

Methods: lexicon-based (gazeteers)
grammar-based (rule-based)
statistical models (machine learning)
hybrids
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

Methods: lexicon-based (gazetteers)
         grammar-based (rule-based)
         ✓ statistical models (machine learning)
         ✓ hybrids
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

Who? Where?

When?
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

**Word Sense Disambiguation**: identifying which sense/meaning of a word is used in a sentence, when the word has multiple meanings.

**Text normalization**: transforming text into a single canonical form that it might not have had before.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

Sam Arlington initiated partnership discussions during his visit to Eureka offices in September.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

J. Smith went to Washington DC to see the Smithsonian Institute and also met up with Virginia Peterson for a coffee.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

Sam Arlington initiated partnership discussions during his visit to Eureka offices in September.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

J. Smith went to Washington DC to see the Smithsonian Institute and also met up with Virginia Peterson for a coffee.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

What?
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

How? Why? How do we feel about it?
S. Arlington initiated partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and also met up with Virginia for a coffee.

How? Why? How do we feel about it?

S. Arlington visited the Eureka’s Ltd offices last month to initiate partnership discussions.

John Smith was delighted to go to Washington to see the Smithsonian and also met up with Virginia for a coffee.
Common problems

Name boundaries... Disambiguation... Normalization...

Also:

S. Arlington did not initiate partnership discussions during his visit to Eureka’s Ltd offices last month.

John Smith went to Washington to see the Smithsonian and almost also met up with Virginia for a coffee.

Negation... Hedging... Conditional phrases...
Eye drops off shelf.

Include your children when baking cookies.

Turn right here.

John saw the man on the mountain with a telescope.

He gave her cat food.

They are hunting dogs.

"Make my speech so ambiguous it doesn't matter what they take out of context!"
Looking for: interactions between SAF and viral LTR elements
(SAF is a transcription factor, LTR stands for ‘long terminal repeat’)

Gene names:
tinman, lilliputian, dreadlocks, lush, cheap date, methuselah, Van Gogh, maggie, brainiac, grim, reaper, cleopatra, swiss cheese, ken and barbie, kenny, out cold, lava lamp, hamlet, sonic hedgehog, werewolf, half pint, fucK, drop dead, chardonnay, agnostic, I’m not dead yet...
Applications

Knowledge management - Extracting metadata

Search engines

Question Answering

Text mining / knowledge discovery

Analytics & trend analysis
• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell

• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features

• Take away message
Specialized domain study:

1. Search interface features: Biomedical Scientists (HCIR)

Anna Divoli and Alyona Medelyan
Search interface feature evaluation in biosciences, HCIR 2011, Google, Mountain View, CA

UI preferred features studies:

2. Study existing popular systems (EuroHCIR)

Matthew Pike, Max L. Wilson, Anna Divoli and Alyona Medelyan
CUES: Cognitive Usability Evaluation System, EuroHCIR 2012, Nijmegen, Netherlands

3. Mock ups of specific features (survey)
1. Search interface features: Biomedical Scientists

- **autocomplete/autosuggest**: predicting a word or phrase that the user wants to type in without the user actually typing it in completely.
- **search/query expansion**: reformulating a seed query to improve retrieval performance.
- **faceted refinement**: information represented using a faceted classification, allowing users to explore by filtering available information.
- **related searches**: suggestions aiming to help users find out common phrases that other users have searched for.
- **results preview**: combination of title, author names, dates, sources, excerpts etc to help users evaluate the pertinence of the result.
Facets – favorite feature for search systems

Anna Divoli and Alyona Medelyan, Search interface feature evaluation in biosciences, HCIR 2011, Google, Mountain View, CA, USA
Facets (in search systems)

<table>
<thead>
<tr>
<th>Molecule Role</th>
<th>Molecule Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>peptide</td>
<td>peptide</td>
</tr>
<tr>
<td>protein</td>
<td>protein</td>
</tr>
<tr>
<td>enzyme</td>
<td>enzyme</td>
</tr>
<tr>
<td>base</td>
<td>base</td>
</tr>
<tr>
<td>gamma-aminobutyric acid</td>
<td>gamma-aminobutyric acid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>dementia</td>
<td>dementia</td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>Prion</td>
<td>Prion</td>
</tr>
<tr>
<td>Huntington's disease</td>
<td>Huntington's disease</td>
</tr>
<tr>
<td>Dystrophy</td>
<td>Dystrophy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cellular component</th>
<th>Cellular component</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell</td>
<td>cell</td>
</tr>
<tr>
<td>host</td>
<td>host</td>
</tr>
<tr>
<td>immunoglobulin complex, circulating</td>
<td>immunoglobulin complex, circulating</td>
</tr>
<tr>
<td>insoluble fraction</td>
<td>insoluble fraction</td>
</tr>
<tr>
<td>pore complex</td>
<td>pore complex</td>
</tr>
</tbody>
</table>

animal models huntington disease
animal models huntington disease
Faceted search is the *most important stand alone feature* in a search interface for bioscientists.

Few, query-oriented facets presented as *checkboxes work best*.

Overly *simple aesthetics*, although not desirable, do not hurt overall UI score.

*Complex aesthetics* turn users away from the systems.

Bioscientists prefer tools that help them *narrow their search*, not expand it.

For *generic search*: *doc-based* facets.
For *domain-specific search*: *query-based* facets.
Facets as search feature: likes & dislikes

**Facetted refinement**

- Useful categories
- Simple
- Vertical list

- Too complex/busy
- Too many colors
- Poor design
- Limited functionality
- Too many symbols
- Not special/ Colorless
Outline

• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell

• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features

• Take away message
Exploring UI features - Systems Tested: Yippy, Carrot, MeSH, ESD

Anna Divoli  CS4HS 2013 @ Unitec
Exploring UI features - Systems Tested: Yippy, Carrot, MeSH, ESD
Exploring UI features - Systems Tested: Yippy, Carrot, MeSH, ESD

Education [102]
Education, Nonprofessional [02.233]
Education, Special [02.233.213] +
Health Education [02.233.332] +
Physical Education and Training [02.233.543] +
Schools [02.233.659]
Schools, Nursery [02.233.659.660]
Students [02.233.748] +
Vocational Education [02.233.862]

Education [102]
Curriculum [02.158] +
Education, Distance [02.195]
Education, Nonprofessional [02.233] +
Education, Predental [02.275]
Education, Premedical [02.316]
Education, Professional [02.358] +
Educational Measurement [02.399] +
Faculty [02.463] +
Inservice Training [02.594] +
International Educational Exchange [02.581]
Mentors [02.588]
Needs Assessment [02.594]
Preceptorship [02.660]

Schools [02.783]
Library Schools [02.783.338]
Schools, Health Occupations [02.783.660] +
Universities [02.783.850]

Students [02.851] +
Teaching [02.901] +
Exploring UI features - Systems Tested: Yippy, Carrot, MeSH, ESD
Exploring UI features - Systems Tested: Yippy, Carrot, MeSH, ESD
Exploring UI features (Yippy, Carrot, MeSH, ESD): likes & dislikes

- Menu highlighting
- Hierarchical folder layout
- Expand hierarchy with “+” and “−”
- Dual view (tree on left, results on right)
- Ability to change visualisations of taxonomy
- Search function is important
- Familiar interface with folders

- Too simple or too much writing - would be nice to have color
- Lots of scrolling
- Dots in carrot circle – confusing
- Double click on foam tree is unintuitive
- Too broad taxonomies

Anna Divoli  CS4HS 2013 @ Unitec
Outline

• Familiar Interfaces
• User Centered Design & Usability
• Who am I & who is Pingar
• Text analytics in a nutshell

• Our usability studies & findings
  • 1. Search interface features: Biomedical Scientists
  • 2. Study existing popular systems
  • 3. Mock ups of specific features

• Take away message

Anna Divoli  CS4HS 2013 @ Unitec
Taxonomy UI preferences: The 51 participants

Age:
- 25 or younger: 27.3%
- 26-40: 60.0%
- 41-60: 12.7%
- 61 or older: 0%

Highest level of education:
- High School: 3.6%
- College/University: 52.7%
- Graduate School: 43.6%

How comfortable you are with computers?
- Somewhat: 5.5%
- Very: 47.3%
- Second nature: 47.3%

Do you have experience using taxonomies?
- No: 30.9%
- Yes, but very little: 47.3%
- Yes: 21.8%
Concept sorting

- A: Sedans (10), Coupes (7), Vans (5), SUVs (3), Trucks (2)
- B: Coupes (7), Sedans (10), SUVs (3), Trucks (2), Vans (5)

- Popularity (A): 44.2%
- Alphabetically (B): 42.3%
- No preference: 13.5%
Displaying Counts

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cars (5)</td>
<td></td>
</tr>
<tr>
<td>Sedans (10)</td>
<td></td>
</tr>
<tr>
<td>Coupes (7)</td>
<td></td>
</tr>
<tr>
<td>Vans (5)</td>
<td></td>
</tr>
<tr>
<td>SUVs (3)</td>
<td></td>
</tr>
<tr>
<td>Trucks (2)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cars (5)</td>
<td></td>
</tr>
<tr>
<td>Coupes (2)</td>
<td></td>
</tr>
<tr>
<td>Sedans (2)</td>
<td></td>
</tr>
<tr>
<td>SUVs (1)</td>
<td></td>
</tr>
</tbody>
</table>

A: 42.3%
B: 51.9%
No preference: 5.8%
Using Labels

- **Transport**
  - **Cars**
    - Sedans
    - Coupes
    - Vans
    - SUVs
    - Trucks
  - Bikes
  - Trains

- **Transport**
  - **Cars**
    - Sedans
    - Coupes
    - Vans
    - SUVs
    - Trucks
  - Bikes
  - Trains

- in frames (A): 72.5%
- with labels (B): 23.5%
- no preference: 3.9%
Plus/minus signs or arrows

A

- Transport
  - Cars
    - Sedans
    - Coupes
    - Vans
    - SUVs
    - Trucks
  - Bikes
  - Trains

B

- Transport
  - Cars
    - Sedans
    - Coupes
    - Vans
    - SUVs
    - Trucks
  - Bikes
  - Trains

A: 47.1%
B: 37.3%
no preference: 15.7%
Search Functionality

- **partial** 74.5%
- **hidden** 64.7%
- **no preference** 2.0%

**Chevrolet**
- **exact match**
- **partial match**

**Chevrolet Aveo LT**
- **partial match**

**Impala LS**
- **hidden match (implies “Chevrolet Impala LS”)**
Outcome: User-friendly Taxonomy Editor

Upload a Taxonomy

You can upload any taxonomy into the Pingar Taxonomy Editor. Just browse, select and upload. Simple!

Create a new Taxonomy

The Pingar Taxonomy Editor allows you to start a taxonomy from scratch. Simply enter a name and begin.

View, Navigate, Edit & Save

The Pingar Taxonomy Editor offers Tree and A-Z views for navigation. You may review every concept’s label and relationships on the right hand side panel. Just select the node you’d like to review. Editing is available on a right click menu or just drag and drop to move the nodes around. Search and edit history are also available. Don’t forget to save your edits!

Learn more...

It is new simple to create and update taxonomies. Review and edit concepts and relationships. To find out more about the Pingar Taxonomy Editor click below:

Learn More

Supported Formats

The Taxonomy Editor v1.1.0 supports 3 different formats: SKOS, XML and CSV. Please be aware that not all formats support all taxonomy features, which may lead to data loss. Both SKOS and XML can be directly used with the Pingar API.

SKOS: An RDF format promoted by W3C for storing knowledge structures. SKOS supports all features.

XML: An EDG UK format which is more readable than SKOS. XML does not support different languages.

CSV: A format used by SharePoint’s Team Manager. CSV does not support different languages, only allows one root concept and up to 7 nesting levels.
Outcome: User-friendly Taxonomy Editor

- Morality (2)
  - Crime (27)
  - Torture (1)
- New England Association of Schools and Colleges (2)
  - Harvard University (2)
  - Massachusetts Institute of Technology
- New Zealand long-distance runners (2)
  - Arthur Lydiard
  - Murray Halberg
- New Zealand middle distance runners (2)
- Novels first (9)
  - Bleak House
  - Dombey and Son
  - Great Expectations
  - Martin Chuzzlewit
  - Nicholas Nickleby
  - Oliver Twist
  - Our Mutual Friend
  - The Brothers Karamazov
  - The Pickwick Papers
- Nuclear weapon (1)
  - Nuclear warfare (1)
- People (191)
- Science, technology and innovation (9)
  - Engineering (2)
  - Microbiology (2)
  - Physics (13)
  - Psychology (2)
  - Research (1)
  - Science (11)
  - Atomic mass unit
  - Biochemistry (18)
  - Biology (20)
    - Animal testing (2)
    - Bacteria
    - Biotechnology (12)
    - Cell (biology)
    - Cell cycle (3)
    - Cell imaging (2)
    - Cell signaling (7)
      - Apoptosis (3)
      - Ligand (biochemistry)
      - Phospholipase
      - Phosphorylation (1)
      - Receptor (biochemistry)
Lessons Learnt

- Follow basic design principles: alignments, right use of colors, etc.
- Avoid clutter! Go for clean and simple always!

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.

*Antoine de Saint-Exupery, French writer (1900 - 1944)*

- Aim for intuitive User Interfaces. Do not make users think!
- You will never please everybody! Choose the users and use cases you plan to support.
Acknowledgements: Pingar Research Team

Anna Divoli
PhD (2006) in Biomedical Text Mining, University of Manchester
Previously: University of Chicago, UC Berkeley, Inpharmatica
Expert in: biomedical text mining, text analytics, user search interfaces, usability

Ignas Kukenys
PhD (2010) in Computer Vision, University of Otago
Previously: NextWindow
Expert in: computer vision, machine learning, classification methods

Alyona Medelyan
PhD (2009) in Natural Language Processing, University of Waikato
Previously: Google, European Media Lab
Expert in: keyword and entity extraction, Wikipedia mining

Ralf Heese
Diploma in Computer Science, PhD candidate, Humboldt-Universität Berlin
Previously: Humboldt-Universität Berlin, Freie Universität Berlin
Expert in: semantic databases, semantic technologies

Hayden Curran
Computer Science, University of Waikato (2008-2010)

Andy Chao
BSc (2011) in Computer and Information Sciences, AUT

Collaborators
Annika Hinze
Waikato Uni

Ian Witten
Waikato Uni

Max Wilson
Nottingham Uni

Jeen Broekstra
Rivuli Development

Alumni
Matthew Pike
Nottingham Uni PhDc

Steve Manion
Canterbury Uni PhDc

David Milne
CSIRO postdoc

Jose Chen
NIH fellow

Anna Huang
Yangtze Uni Lecturer

And all 65+ anonymous studies participants!
Take away message: Involve the user!

By Frits Ahlefeldt

http://farm5.static.flickr.com/4087/5019944289_5e2f0637c9.jpg