Seeking Health on the Web

ICML 2009

Tamas Doszkocs, Ph.D.
Computer Scientist
doszkocs@nlm.nih.gov
Sex... and ... Health

After sex, health is one of the most searched topics on the Web

- 61% of American adults look online for health information
- 86% of all adults also ask a health professional
- 68% of all adults ask a friend or family member
- 54% use books or other printed reference material.
- A majority of e-patients access user-generated health information
- But few are actively writing or creating new health content
- Online health inquiries have an impact on decisions or actions and there are clearly more positive experiences than negative ones

The Social Life of Health Information June 2009
Now that I have your attention 😊

“Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the MEMEX and there amplified”

Vannevar Bush “As We May Think”

JULY 1945 ATLANTIC
Trends in Searching 2008-2009

- The Web as the MEMEX
- The Semantic Web
- Mashups and RIA
- Vertical Search
- Universal Search
- Semantic Search
- Discovery Search

- Connected Mobility
- Focus on Consumers
- Information Democracy
- Social content
- Social search
  - Social evaluation
- Information Monopolies
- Dominance of Keyword Wizardry
Utilization of Natural Language Processing Techniques

• Behind the scenes
  • Statistical NLP
  • Linguistic NLP
  • AI Heuristics

• Innovation and Competition
  • The BIG 3

• Sample query: Why Me?
  • Ask Q&A
  • Bing
  • Google Squared
  • Yahoo!
Trends in Health Information

• Health Care Reform
• New Health Resources
• Always-on Connections
• Personally Relevant Information
• The Evolving Semantic Web
• New or Improved Health Search Engines
• Semantic Health Search

• Participatory Medicine
• E-patients
• first, second and third opinions on and off the Web
  – Health Professionals
  – Friends
  – Family

• Informed Personal Health Decisions
Consumer Health Search Engines

high quality health content

• http://Medlineplus.gov
• http://MayoClinic.com
• http://WebMD.com
  • http://www.uptodate.com/patients/index.html
  • http://www.intelihealth.com
  • http://www.lib.uiowa.edu/hardin/md/
• http://healthfinder.gov/

• http://medem.com
• http://www.medhelp.org/
• http://www.medicinenet.com/
• http://www.merck.com/consumer/
• http://my.clevelandclinic.org/
• http://www.hon.ch/MedHunt/
• many others
Newer Health Search Engines
trustworthy health information

• http://www.healia.com/
• http://medstory.com
• http://mednar.com/
• http://www.searchmedica.com/
• http://www.netwellness.org/
• http://www.nextbio.com/
• http://righthealth.com/
• http://www.scienceroll.com
• http://HealthoPedia.com
• http://hso.info/
References

- [http://www.mlanet.org/resources/consumr_index.html](http://www.mlanet.org/resources/consumr_index.html)
- [http://caphis.mlanet.org/consumer/index.html](http://caphis.mlanet.org/consumer/index.html)
- The Engaged E-patient Population 8/26/2008
- Mobile Could Be a Game Changer
- “Seeking health information online: does Wikipedia matter?”, Laurent MR, Vickers TJ., J Am Med Inform Assoc. 2009 Apr 23,
  - “Wikipedia surpassed MedlinePlus and NHS Direct Online … and ranked higher with quality articles”
- Challenges and Promises of the Semantic Web Olivier Bodenreider, SLA 2009
- The Social Life of Health Information June 2009
Semantic Search

Goal: better and more contextual search results

- SemanticSearch.org
- http://www.searchenginejournal.com/askcom-focuses-on-semantic-search/8252/
- http://www.powerset.com/
- http://www.hakia.com/
- http://www.newssift.com
Health Search Engines with Semantic Search Capabilities

combine human expertise and semantic knowledge

• **HealthLine**
  - uses own taxonomy of > 250,000 health terms
  - thousands of Indian doctors and pharmacists
  - [http://health.msn.com](http://health.msn.com)


• [http://www.semanticmedline.com/](http://www.semanticmedline.com/)
HealthMash
a semantic health search and discovery engine

- [http://healthmash.com/](http://healthmash.com/) is an innovative next generation semantic health search engine
- HealthMash developers have been working on NLM’s ToxSeek and AidsSeek projects for over 5 years
- Beta version of HealthMash was showcased at MLA 2009
- HealthMash utilizes a pragmatic mix of natural language processing tools, semantic engineering techniques and multiple knowledge sources, including a proprietary **Health Knowledge Base**, to achieve both high precision and relevancy in its search results
“The HealthMash Premise”

trusted health information is vital for

• dealing with health problems,

• promoting healthy behavior,

• making healthy decisions and

• for overall well being

– However, as Mark Twain put it: Be careful of reading health books. You may die of a misprint! (1835 - 1910)
“The HealthMash Promise”

the information must be

– **Reliable**, 

– **Relevant**, 

– **Recent** and 

– **Related** to the search topic
Revolutionary Health Knowledge Base and Semantic Search Engine
Query: bipolar disorder

- The Result Page consists of **FOUR** types of information:
  - **The Search Results** from trusted sources (the middle column) are retrieved by our vertical semantic search engine to produce **RELIABLE** information
  - **The Meta-Search Results** (the 3rd column) show **RECENT** News, Video and other multi-dimensional information…
  - **The Combined Table of Contents** from the trusted sources (left column) shows **RELEVANT** content links that allow the user to drill-down in the search results
  - **The Explore and Discover Table** presents specific **RELATED** information that is closely associated with the query at hand and allows the user to meaningfully and dynamically **shift focus** and/or **MODIFY** the search strategy
The Explore and Discover data comes the Health Knowledge Base

- The *Health Knowledge Base* is automatically generated from trusted health content sites and diverse knowledge sources, such as MeSH and UMLS. HealthMash also utilizes the Web itself as a data base.

- The *Health Knowledge Base* contains explicit knowledge about
  - Health Concerns
  - Causes
  - Signs and Symptoms
  - Tests, Procedures, Treatments
  - Drugs and Substances and adverse effects
  - Alternative, Complementary and Integrative Medicine

- The *Health Knowledge Base* is also available via a web service

- The *Health Knowledge Base* facilitates exploration and discovery
HealthMash reflects the most notable trends in searching in 2009, namely

- **Vertical Search**
- **Universal Search**
- **Semantic Search**
  - Semantic Search attempts to understand the MEANING of Web content and search queries
- HealthMash combines all of these search techniques
Query: **gout**

- As you can see in this typical search example, the **SearchResults** in HealthMash are quality retrievals from **Reliable** sources, such as *Medlineplus*, *MayoClinic* etc.

- The **Health News, Blogs, Twitter** etc are **Recent** and often **real time** information.

- The **Contents** links lead to **Relevant** information on **causes, complications** etc, and the **mouse-overs** show detailed **CONTEXT**.

- The **Explore and Discover** table lists specific **Drugs** (*colchicine, allopurinol* etc.) and other important information (**Health Concerns, Treatments** etc.) that is **specifically Related** to **gout** (and here too the **mouse-overs** show sufficient **CONTEXT**).
Characteristics of a Useful Semantic Health Search Engine

• The search engine must be as good or better than a group of experts in identifying reliable information sources and analyzing and synthesizing relevant findings,

• Must present the results in a clearly organized manner,

• Must provide sufficient CONTEXT for the user,

• Must offer “second opinions”, or better yet, “multiple opinions”

• Must provide answers to unasked questions and

• Must facilitate good choices and decisions
Query: **heavy drinking**

- Not all queries have data in the Health Knowledge Base
- In such situations HealthMash performs **dynamic faceted clustering** of the search results in order to support focused drill-down
The **Semantic Query and Knowledge Base Processing in HealthMash** is accomplished by

- Proprietary [Natural Language Processing](#) tools, such as
  - Lexical/morphological and orthographic tools,
  - Syntactic tools, and
  - Semantic tools

- HealthMash also utilizes *semantic knowledge sources*, such as
  - [MESH](#) (the Medical Subject Headings Thesaurus of the NLM),
  - [UMLS](#) (the Unified Medical Language System of NLM/NIH) and other semantic data repositories and
  - The Health Knowledge Base

- The **Health Knowledge Base** is the most important *semantic resource* in HealthMash
HealthMash

• extensively utilizes **Web 2.0 techniques**
  – Asynchronous information processing
  – Mashups
  – Social networking (under development)

• It is also important to note that the solutions and tools used in HealthMash are, for the most part, **domain independent**, and as such they are equally applicable to Energy, Law, Education and other areas of human endeavor.
From a tools perspective, HealthMash consists of the following components:

- **PolyDictionary** (Medical and Scientific Dictionary System)
- **PolySpell** (medical and scientific spell checkers)
- **PolyTagger** (part-of-speech tagger)
- **PolyPhraser** (noun phrase parser)
- **PolySearch** (intelligent concept search engine)

  » Query: inflamed testicles
- **PolyCluster** (search result clustering engine)
- **PolyMeta** (federated search and discovery engine)
From a technical perspective, the **Explore and Discover** table shows important health related entities, identified by a mix of

- **Linguistic Engineering** and

- **Statistical Techniques**, as well as

- **Heuristics**, using WebLib’s

- **Proprietary Semantic Health Knowledge Base**

- **Proprietary Semantic Search Algorithms** and

- **Proprietary Semantic Ranking** techniques.
IN SUMMARY

– the automatically generated and automatically enhanced Health Knowledge Base is the key value-added semantic component of HealthMash

– the proprietary semantic search and semantic ranking technologies utilized in HealthMash contribute to better quality search results
NIH Library Search Demo

NIH Library AllPlus Search Demo

Default Search Evidence-based Medicine Academic Web Alternative Medicine Animal Studies Drugs and Chemicals Search Engines News Video Images Consumer Health

Keyword ▼ no periods Search Select Sources
The NIH Library

- Collaboration between NLM’s Specialized Information Services Division and the NIH Library
- Utilizes WebLib’s PolyMeta Federated Search and Clustering Engine
- Will incorporate WebLib’s HealthMash Semantic Search Engine and Health Knowledge Base
Query Processing

- User queries may imply multiple concepts, e.g.
  - “Abdominal pain in children”
  - “diet and exercise”
- Queries may be ambiguous, for instance
  - “No periods”
- Queries are often misspelled, e.g.
  - Diareaa
Please take good care of yourselves and remember that

It's no longer a question of staying healthy.

It's a question of finding a sickness you like.

Jackie Mason (1934 - )
American comedian
Seeking Health on the Web

ICML 2009

Tamas Doszkocs, Ph.D.
Computer Scientist
doszkocs@nlm.nih.gov