Using Ontologies to Create a Flexible Search Option

Brandy E. King, MLIS
Center on Media and Child Health (CMCH)

Children’s Hospital Boston,
Harvard Medical School, Harvard School of Public Health
The Need for CMCH

By the year 2000, the American child between the ages of 8 and 18 years spent 6 hours and 43 minutes of every day with media, more time than they spent in school or with parents.
The Need for CMCH

Attention to the subject from medical organizations:

- American Medical Association

- American Academy of Pediatrics policy statements
  - Impact of Music (1995)
  - Violence (2001)
  - Substance abuse (2001)
  - Risky sexual behavior (2001)
  - Obesity (2003)
The Need for CMCH

CMCH Mission:
To advance scientific research, clinical interventions, and education on the subject of media and their effects, positive and negative, on the physical and mental health of children and adolescents.
CMCH Database of Research

One database to bring together all existing research on the effects of media on child health

Searchable by researchers from any discipline

Searchable by novice users from the general public
Challenges

1. Multiple Disciplines
2. Multiple Vocabularies
3. Multiple Audiences with varying levels of search experience
Challenges:

Multidisciplinary ➔ Interdisciplinary

Media effects research is currently multidisciplinary:

- medicine
- psychology
- public health
- education
- gender studies
- communication
- sociology
- anthropology
- business
- criminal justice
Challenges:
Ambiguous Vocabulary

Example -

- Aggression
- Aggressiveness
- Aggressive Behavior

Has a negative connotation in almost every field except psychology, where it can also have the positive meaning of a self-confident personality.
Challenges:

**Ambiguous Vocabulary**

Example

Comparison of Search Results from Thumbshots.com

<table>
<thead>
<tr>
<th>Search Terms</th>
<th>Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movies vs. Films</td>
<td>5%</td>
</tr>
<tr>
<td>Media vs. “Mass Media”</td>
<td>3%</td>
</tr>
<tr>
<td>Video Games vs. Electronic Games</td>
<td>2%</td>
</tr>
<tr>
<td>Video Games vs. Computer Games</td>
<td>27%</td>
</tr>
<tr>
<td>Internet vs. “World Wide Web”</td>
<td>0%</td>
</tr>
</tbody>
</table>
Challenges:
Multiple Audiences

At least 6 different audiences:

- Researchers
- Clinicians
- Parents
- Teachers
- Policymakers
- Children and Adolescents
Searching: CMCH Database of Research

Needed a search mechanism to translate across disciplines, across vocabularies, and across audiences

Semantic Search of CMCH Database
Solutions:
Discovering Ontologies

SemanTx Life Sciences Company

Semantic Knowledge Indexing Platform (SKIP)

Runs on National Library of Medicine ontology: Unified Medical Language System (UMLS)
Solutions:
Discovering Ontologies

Ontology:
An approved vocabulary containing defined concepts and the describable relationships between them
Discovering Ontologies

Most common relationship

Is_A = is a type of

Human Is_A Mammal

153 possible relationships:

brought_about_by         diagnosed_by
has_ingredient           result_of
spatially_related_to     traversed_by
Comparison of Search Engines

Entity Extraction
- Taxonomy
- Syntactic
- Directory
- Pattern Match
- Cluster
- Word Match/Key Words

Idea Extraction
- Ontologies

Semantic Search

- Clearforest
- Verity, Convera, Endeca
- iPhrase
- Vivisimo
- Google

Quigo (categorization)
- Inxight, Fast

Autonomy

MSN
Discovering Ontologies

Unified Medical Language System (UMLS)

- Created by the National Library of Medicine
- Contains over 1 million concepts
- 5 million unique concept names
- Created from more than 100 different source vocabularies
Creating the CMCH Ontology

1. Created new concepts and definitions if the terms were not part of the UMLS already

Mean World Syndrome: theory that the more television people watch, the more they believe that violence is a commonplace as it is on television.

Third Person Effect: the belief that media can and does affect the way people think, but that you are personally exempt from these effects.
Creating the CMCH Ontology

2. Added synonyms to reflect the many variations of terms

**Motion Pictures**
Other Names: Movies
Film
Cinema

**Imitation**
Other Names: Modeling
Copying
Observational Learning
Creating the CMCH Ontology

3. Created new categories

- **Entity**
  - Conceptual Entity
  - Intellectual Product

- **Media Forms:**
  - Advertising
  - Animation
  - Fashion
  - Media Campaigns
  - News
  - Pornography
  - Public Service Announcement

- **Media Types:**
  - Audio Media: Radio, Rap Music
  - Computer and Video Game Media: Internet, Playstation
  - Film Media: Motion Pictures, Videotapes
  - Print Media: Books, Magazines
  - Television Media: Cartoons, Children’s Television

- **Media Content:**
  - Aggression (Media Content)
  - Alcohol (Media Content)
  - Bullying (Media Content)
  - Cigarettes (Media Content)
  - Crime (Media Content)
  - Homicide (Media Concept)
  - Marijuana (Media Content)
  - Smoking (Media Content)
  - Steroids (Media Content)
  - Tobacco (Media Content)
  - Violence (Media Content)
  - Weapons (Media Content)
Creating the CMCH Ontology

4. Assigned concepts to different categories where necessary
Creating the CMCH Ontology

5. Added relationships to enhance meaning

Example: Instant Messaging

<table>
<thead>
<tr>
<th>Semantic Relationship</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is_A</td>
<td>Communications</td>
</tr>
<tr>
<td>Occurs_In</td>
<td>Internet</td>
</tr>
<tr>
<td>Occurs_In</td>
<td>Computers</td>
</tr>
<tr>
<td>Performed_By</td>
<td>Peer Group</td>
</tr>
<tr>
<td>Has_Occurrence</td>
<td>Interpersonal Communication</td>
</tr>
<tr>
<td>Has_Occurrence</td>
<td>Bullying</td>
</tr>
</tbody>
</table>
Searching:
CMCH Database of Research

CMCH Smart Search:
Translates queries across disciplines, across vocabularies, and across audiences

Semantic Search of CMCH Database
Comparison of Search Engines

- Semantic Search
- Idea Extraction
- Ontologies
- Entity Extraction
- Taxonomy
- Syntactic
- Directory
- Pattern Match
- Cluster
- Word Match/Key Words

Search Engines:
- Clearforest
- Quigo (categorization)
- Inxight, Fast
- Autonomy
- Verity, Convera, Endeca
- iPhrase
- Vivisimo
- Google
- MSN
Contact Information

Brandy King
brandy.king@childrens.harvard.edu

Center on Media and Child Health
www.cmch.tv