E-Health: How Semantic Retrieval Benefits Patients

or

Automatic Learning of Lightweight Ontologies

Ulrich Reimer

ai-one

&

Institute for Information and Process Management
University of Applied Sciences St. Gallen
Switzerland
Outline

1. What SEMPER is about
2. The need for semantic search
3. Learning lightweight ontologies with biologically inspired neural nets
4. Evaluation of learning lightweight ontologies
5. Future Work: Fact retrieval
6. Outlook
SEMPER: Fact Sheet

Motivation:
• increase in chronic diseases causes growing health costs
• support and education for self-care is essential for compliance and long-term behavioural change
• new media have great potential in an integrated health care approach, esp. for personalised monitoring and information access

Application scenarios:
• work-related health problems
• alcohol-related problems

Goals:
• empowering patients
• Web-based platform for patient self-management
Functional Modules of the Self-Management System

- **Decision Support**: Goal setting for personal action plans
- **Motivation Support**: Continuous feedback and monitoring of an agreed upon action plan
- **Social Support**: e.g. through online support groups, discussion forums, blogs
- **Information Portal**: (access to „official“ and informal knowledge)
Motivation and Monitoring Support

Alcohol dependency

• Self-help tool with self-assessment and self-management for anonymous visitors
• Online community support for after-care of ex-patients
• Success stories
Motivation Support

Assessment of beliefs, behaviour, and knowledge regarding self-management

Personal Action Plan
1. Lists specific goals in behavioural terms
2. Lists likely barriers and plans to overcome them
3. Lists follow-up plan
4. Is shared with all members of the health team

Identification of personal barriers and support (e.g. mobility constraints)

Collaborative goal setting

Development of personalised strategies and problem solving
Motivation and Monitoring Support

Work-related disorders

• Ergonomic check of workplace with natural language dialogue
• Compliance by means of group pressure and incentives offered by employer
Community Support

Benefits of virtual support groups:

• Participation independent of time and space → suitable for people living in remote areas or people homebound due to illness, age or handicap
• Anonymity is often desirable for exchange on stigmatised (health) problems
• Online communities are an untapped knowledge resource

Virtual support groups are integrated into the search function
Semantic Information Portal

3 Types of Information Sources:

1. **Internal:**
   high-quality information from organisations and experts, e.g. specific clinical pictures, case histories, activities/actions to reduce or eliminate symptoms

2. **External from community sites:**
   Information gathered from discussion forums, blogs etc.

3. **External from the Web:**
   from relevant information sources on the Web
Outline

1. What SEMPER is about

2. The need for semantic search

3. Learning lightweight ontologies with biologically inspired neural nets

4. Evaluation of learning lightweight ontologies

5. Future Work: Fact retrieval

6. Outlook
Why Semantic Search?

• Huge gap between information needs of patients and their transformation into appropriate queries

• People often don’t know what they are looking for

• Popular terms vs. subject-specific terminology (esp. in the medical domain)
Problems with Standard Text Retrieval (1)

treatment of a sore throat

Are there any home remedies for a sore throat?

... Salt water gargles, hard candies, sprays and lozenges can provide temporary pain relief.

... A humidifier may be helpful in relieving symptoms, especially in sore throats caused by mouth breathing and dry air.

Antibiotics are prescribed for pharyngitis caused by bacteria. These drugs are effective in killing bacteria, and certain other organisms, but not viruses.
Query Extension through Background Knowledge (1)

treatment of a sore throat

full-fledged ontology

sore throat $\text{syn}$ pharyngitis

treatment

drug $\text{syn}$ medicine $\text{is-a}$ gargles $\cdots$ pain relief $\text{is-a}$ spray $\cdots$ lozenge $\text{is-a}$

(treatment OR medicine OR drug OR gargles OR … ) AND
(sore throat OR pharyngitis)
Problems with Standard Text Retrieval (2)

lifestyle “back pain”

... However, if you are overweight or obese, chances are you have, or will have, back pain. ...

... There is a strong connection between stress and back pain. ...

... studies show that there is a direct relationship between smoking cigarettes and having back problems. ...

...
Query Extension through Background Knowledge (2)

(lifestyle OR physical exercise OR stress OR smoking OR obesity OR ...) AND “back pain”
Query Extension through Background Knowledge (3)

(lifestyle OR physical exercise OR stress OR smoking OR obesity OR …) AND “back pain”
Navigate the Search Results

lifestyle “back pain”

automatic query extension

lifestyle
  physical exercise
    Physical exercise against back pain
    Back pain – more daily exercise
    Relieving back pain with yoga
    Back pain: Exercise instead of bed rest

nutrition
smoking
stress
overweight
Lightweight Ontologies: Association Strength

- Physical exercise
  - 0.9
  - Lifestyle
    - 0.75
    - Stress
      - 0.6
      - Obesity
    - 0.6
    - Nutrition
      - 0.9

→ Lightweight ontologies are also called associative networks
Clinical presentations

A 2-month-old boy with scoliosis and a protruding bulge. Most impressive are the tympanic resonance on percussion due to the high air content. There is no vascular disorder to thrive.

Local and general findings

Slightly differential diagnosis shows the broad findings of a more distinct type, it is mainly a transitional type of a broad, mould-like.
Where Do the Ontologies Come From?

Possible solutions:

• manual construction: in most cases too costly or impossible

• ontology learning: difficult for full-fledged ontologies

• sharing / reuse of ontologies: limited possibilities

• folksonomies: more a supplement than a solution
Where Do the Ontologies Come From?

Possible solutions:

• manual construction: in most cases too costly or impossible

• **ontology learning**: we do it for lightweight ontologies

• sharing / reuse of ontologies: limited possibilities

• folksonomies: more a supplement than a solution
Outline

1. What SEMPER is about
2. The need for semantic search
3. Learning lightweight ontologies with biologically inspired neural nets
4. Evaluation of learning lightweight ontologies
5. Future Work: Fact retrieval
6. Outlook
Learning Scenario

Initial learning step

- text documents
- linguistic preprocessing
- chunks of seed concepts
- ai-one™ neural net learning algorithm
- associations between nouns with strength > C

Incremental learning

- query engine
- search results
- multi-term query as chunk of seed concepts

Learning Scenario:

- physical exercise
- lifestyle
- smoking
- stress
- obesity
- nutrition
ai-one™: Biologically Inspired Neural Networks*

Characteristics:

• biologically inspired: neurons with axons, synapses, dendrites
• stimulation by binary spikes, no thresholds
• stimuli are encoded as distribution of spikes over time
• reinforce traversed connections
• weaken non-traversed connections
• create new connections as a reaction to stimuli
• no predefined topology or neighbourhood function

* invented by Manfred Hoffleisch
ai-one™: Biologically Inspired Neural Networks

Providing a notion of relevance by overlapping chunks of concepts:

$chunk_1 = \{ \text{back pain, stress relief, physical exercise, posture, ergonomic workplace} \}$

$chunk_2 = \{ \text{obesity, nutrition, physical exercise, self esteem} \}$
ai-one™: Biologically Inspired Neural Networks

Providing a notion of relevance by overlapping chunks of concepts:

\[ \text{chunk}_1 = \{ \text{back pain, stress relief, physical exercise, posture, ergonomic workplace} \} \]

\[ \text{chunk}_2 = \{ \text{obesity, nutrition, physical exercise, self esteem} \} \]

life philosophy
ai-one™: Biologically Inspired Neural Networks

Providing a notion of relevance by overlapping chunks of concepts:

\[
\text{chunk}_1 = \{ \text{back pain}, \text{stress relief}, \text{physical exercise}, \text{posture}, \text{ergonomic workplace} \}
\]

\[
\text{chunk}_2 = \{ \text{obesity}, \text{nutrition}, \text{physical exercise}, \text{self esteem} \}
\]
ai-one™: Biologically Inspired Neural Networks

Providing a notion of relevance by overlapping chunks of concepts:

\[ \text{chunk}_1 = \{ \text{back pain, stress relief, physical exercise, posture, ergonomic workplace} \} \]

\[ \text{chunk}_2 = \{ \text{obesity, nutrition, physical exercise, self esteem} \} \]

- higher-order term co-occurrences: sub-sentences & chunks
- chunks are hypergraphs
- chunks introduce background knowledge
Outline

1. What SEMPER is about
2. The need for semantic search
3. Learning lightweight ontologies with biologically inspired neural nets
4. Evaluation of learning lightweight ontologies
5. Future Work: Fact retrieval
6. Outlook
Evaluation of Ontology Learning with ai-one™

Comparison with cosine measure from vector space model on a tf-idf document term matrix:

\[ \text{tf}_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \]

\[ \text{idf}_i = \log \frac{|D|}{|\{d \in D | t_i \in d\}|} \]
## Evaluation of Ontology Learning with ai-one™ (1)

<table>
<thead>
<tr>
<th>query term</th>
<th>$c = 0.8$</th>
<th></th>
<th></th>
<th>$c = 0.9$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ai-one™</td>
<td>cos/tf-idf</td>
<td>ai-one™</td>
<td>cos/tf-idf</td>
<td>ai-one™</td>
<td>cos/tf-idf</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>T</td>
<td>R/T</td>
<td>R</td>
<td>T</td>
<td>R/T</td>
</tr>
<tr>
<td>alcohol dependency</td>
<td>42</td>
<td>64</td>
<td>0.66</td>
<td>4</td>
<td>6</td>
<td>0.66</td>
</tr>
<tr>
<td>alcohol problem</td>
<td>44</td>
<td>64</td>
<td>0.69</td>
<td>7</td>
<td>16</td>
<td>0.44</td>
</tr>
<tr>
<td>prevention</td>
<td>13</td>
<td>20</td>
<td>0.65</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>therapy</td>
<td>79</td>
<td>116</td>
<td>0.68</td>
<td>9</td>
<td>13</td>
<td>0.69</td>
</tr>
<tr>
<td>treatment facility</td>
<td>6</td>
<td>8</td>
<td>0.75</td>
<td>10</td>
<td>48</td>
<td>0.21</td>
</tr>
<tr>
<td>withdrawal symptoms</td>
<td>29</td>
<td>32</td>
<td>0.9</td>
<td>15</td>
<td>34</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>average</strong></td>
<td>60.9</td>
<td>0.73</td>
<td>38.3</td>
<td>0.47</td>
<td>46.6</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Evaluation of Ontology Learning with ai-one™ (2)

Further advantages of ai-one™ over other approaches:

• incremental learning
• works already on a small number of input texts
• asymmetric (directed) associations
• computationally tractable
Outline

1. What SEMPER is about
2. The need for semantic search
3. Learning lightweight ontologies with biologically inspired neural nets
4. Evaluation of learning lightweight ontologies
5. Future Work: Fact retrieval
6. Outlook
Hidden Relationships in a Search Formulation

Ibuprofen “side effect”

wanted:
documents about the drug Ibuprofen and the side effects it has

not wanted:
documents where the drug Ibuprofen and the word “side effect” are mentioned
Hidden Relationships in a Search Formulation

Ibuprofen “Morbus Bechterew”

wanted:
documents about the drug Ibuprofen and its use for treating the disease Morbus Bechterew

not wanted:
documents where the drug Ibuprofen and the disease Morbus Bechterew are mentioned
Automatic Ontology Population after a Training Phase

Manually created ontology:

- **drug** → **treatment**
- **part-of**
- **treatment-for**
- **treats**
- **disease**

Documents:

- "Ibuprofen helps with Morbus Bechterew."

AI one → Patterns in the documents
Automatic Ontology Population after a Training Phase

Manually created ontology:

- drug
- treatment-for
- part-of
- treatment
- treats
- disease

"Ibuprofen helps with Morbus Bechterew."

Documents:

- ai one

Patterns in the documents
Automatic Ontology Population after a Training Phase

Manually created ontology:

- **Drug**: Ibuprofen
- **Disease**: Morbus Bechterew
- **Part of**: treatment
- **Treats**: drug

**Pattern in the documents**:

- "Ibuprofen helps with Morbus Bechterew."

Learning general patterns for certain kinds of statements!

→ Learning general patterns for certain kinds of statements!
Automatic Ontology Population after a Training Phase

Manually created ontology:

- **drug** ➔ **treatment**
  - part-of

- **treatment** ➔ **treats** ➔ **disease**

- **Ibuprofen** ➔ **treatment-for** ➔ **osteoarthritis**

- **Ibuprofen** ➔ **treatment-for** ➔ **osteohritis**

- **inst** ➔ **inst**

- 

→ No fact acquisition, only relationships!
Semantic Search

Manually created ontology:

- disease
- drug
- treatment
- Ibuprofen
- osteoarthritis

"Which drugs help with osteoarthritis?"

\[ \Rightarrow \text{No fact retrieval, only document (passage) retrieval!} \]
Outline

1. What SEMPER is about
2. The need for semantic search
3. Learning lightweight ontologies with biologically inspired neural nets
4. Evaluation of learning lightweight ontologies
5. Future Work: Fact retrieval
6. Outlook
Outlook (1)

Semantic Search:

- integrate thesauri/ontologies as further background knowledge
- examine effects of text type, text length, number of texts
- use ai-one™ in supervised learning for information extraction

SEMPER overall:

- address concerns about information quality
- integrate further application partners into the platform
Outlook (2)

Learning lightweight ontologies – further applications:

• creating initial domain models (e.g. cause-effect graphs)
• semantic schema mapping (e.g. for knowledge integration)
• document clustering (e.g. curation of life science databases)
• ...