Keyphrases as Knowledge Units for Text-based Applications

Daqing He
Department of Informatics and Networked Systems
School of Computing and Information
University of Pittsburgh
Outline

• Basics of Keyphrases: Definitions and Importance
• Identification of Keyphrases: Extraction and Generation
• Applications of Keyphrases: knowledge unit for supporting student learning
• Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
• Applications of Keyphrases: knowledge unit for interactive machine learning
• Conclusions
Outline

• Basics of Keyphrases: Definitions and Importance
• Identification of Keyphrases: Extraction and Generation
• Applications of Keyphrases: knowledge unit for supporting student learning
• Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
• Applications of Keyphrases: knowledge unit for interactive machine learning
• Conclusions
Keyphrases: a Definition

Keyphrases: Short noun phrases to summarize and highlight important information in a piece of text

- Examples: see the figure
Keyphrases: a Definition

- Keyphrases: Short noun phrases to summarize and highlight important information in a piece of text
  - Examples: see the figure
Keyphrases: a Definition

▷ Keyphrases: Short noun phrases to summarize and highlight important information in a piece of text
  • Examples: see the figure
Keyphrases: a Definition

▷ Keyphrases: Short noun phrases to summarize and highlight important information in a piece of text
  • Examples: see the figure

▷ Keyphrases are different to words
  • Same as keyphrases, string of characters written by authors
  • Different to keyphrases, individual words might not have a complete and unique meaning
Keyphrases: a Definition

▷ Keyphrases: Short noun phrases to summarize and highlight important information in a piece of text
  • Examples: see the figure

▷ Keyphrases are different to words
  • Same as keyphrases, string of characters written by authors
  • Different to keyphrases, individual words might not have a complete and unique meaning

▷ Keyphrases are different to concepts
  • Same as keyphrases, concepts have a complete and unique meaning, good knowledge unit
  • Different to keyphrases, concepts are more knowledge focus, less language focus, thus need manual construct
Why Keyphrases?

- Keyphrases are a natural and neat way to express important information ➔ semantic and knowledge unit
  - Better than words
- Keyphrases are a natural & efficient language units connecting human and data/information
  - Better than concepts, which need abstraction
- Modern representations of keyphrases enable wider applications of keyphrases
  - Embedding-based representations enable direct computation on keyphrases
  - Enable keyphrases act as knowledge unit, not just text level

Takeaway message
- Keyphrases can combine the benefits of words and concepts in various real world applications
Applications of Keyphrases

- Information retrieval (indexing term)

Applications of Keyphrases

- Information retrieval (indexing term)
- Summarization (locate key sentences)

Applications of Keyphrases

- Information retrieval (indexing term)
- Summarization (locate key sentences)
- Online Advertising

Applications of Keyphrases

▷ Information retrieval (indexing term)
▷ Summarization (locate key sentences)
▷ Online Advertising
▷ Many other applications
Outline

• Basics of Keyphrases: Definitions and Importance
• Identification of Keyphrases: Extraction and Generation
• Applications of Keyphrases: knowledge unit for supporting student learning
• Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
• Applications of Keyphrases: knowledge unit for interactive machine learning
• Conclusions
How to Obtain Keyphrases?

▷ Certainly not manual methods ==> avoid same limitation of concepts
▷ Ideal approaches should be automatic
Digital Texts do Help!

Method 1: Keyphrase Extraction

- Select important words/phrases from the source text
  - Step 1: Generate candidates
  - Step 2: Rank candidates and return top K as results

Automatic Keyphrase Extraction in Textbooks

Framework for manual concept annotation

The meaning of the term information retrieval can be very broad. Just getting a credit card out of your wallet so that you can type in the card number is a form of information retrieval. However, as an academic field of study, information retrieval might be defined thus:

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).

As defined in this way, information retrieval used as an activity that only a few people engaged in: reference librarians, paralegals, and similar professional searchers. Now the world has changed, and hundreds of millions of people engage in information retrieval every day when they use a web search engine or search their email. Information retrieval is fast becoming the dominant form of information access, overtaking traditional database-style searching (the sort that is going on when a clerk says to you: “I’m sorry, I can only look up your order if you can give me your Order ID”).

IR can also cover other kinds of data and information problems beyond that specified in the core definition above. The term “unstructured data” refers to data which does not have clear, semantically overt, easy-for-a-computer structure. It is the opposite of structured data, the canonical example of which is a relational database, of the sort computers usually use to maintain product inventories and personnel records. In reality, almost no data are truly “unstructured”. This is definitely true of all text data if you count the latent linguistic structure of human languages. But even accepting that the intended domain of structure is overt structure, most text has structure, such as headings and paragraphs and footnotes, which is commonly represented in documents by explicit markup (such as the coding underlying web

---

**Table 1** Coding schema for concept annotation

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description (bold text) with Examples and Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (Round 1)</td>
<td>Only noun/noun phrases are considered</td>
</tr>
<tr>
<td>2. (Round 1)</td>
<td>Abbreviation of a concept is also a concept</td>
</tr>
<tr>
<td>3. (Round 1)</td>
<td>Annotate the whole noun/noun phrases, but ignore the general adj. (e.g., long, big etc.)</td>
</tr>
<tr>
<td>4. (Round 2)</td>
<td>If two noun phrases are concepts, the combination should be the concept</td>
</tr>
<tr>
<td>5. (Round 3)</td>
<td>The concepts combined with conjunctions should be separated (e.g., and, or)</td>
</tr>
<tr>
<td>6. (Round 5)</td>
<td>All variations of the concepts should be annotated</td>
</tr>
<tr>
<td>7. (Round 6)</td>
<td>Annotate all special / not general phrases in the</td>
</tr>
<tr>
<td>8. (Round 6)</td>
<td>Ignore the Abbreviation in brackets</td>
</tr>
<tr>
<td>9. (Round 9)</td>
<td>If the concept term has punctuations, keep them</td>
</tr>
<tr>
<td>10. (Round 9)</td>
<td>The well-known and important examples should be annotated</td>
</tr>
</tbody>
</table>

Dataset

- Section-level concept index for the first 16 chapters of the book *Introduction to Information Retrieval (IIR)*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of chapters</td>
<td>16</td>
</tr>
<tr>
<td>Number of sections</td>
<td>86</td>
</tr>
<tr>
<td>Number of all concepts</td>
<td>3175</td>
</tr>
<tr>
<td>Number of 1-grams</td>
<td>1121 (35.31%)</td>
</tr>
<tr>
<td>Number of 2-grams</td>
<td>1565 (49.29%)</td>
</tr>
<tr>
<td>Number of 3-grams</td>
<td>422 (13.29%)</td>
</tr>
<tr>
<td>Number of 4-grams</td>
<td>58 (1.83%)</td>
</tr>
<tr>
<td>Number of 5+6-grams</td>
<td>9 (0.28%)</td>
</tr>
<tr>
<td>Number of all unique concepts</td>
<td>1543</td>
</tr>
<tr>
<td>Number of unique 1-grams</td>
<td>278 (18.02%)</td>
</tr>
<tr>
<td>Number of unique 2-grams</td>
<td>871 (56.45%)</td>
</tr>
<tr>
<td>Number of unique 3-grams</td>
<td>330 (21.39%)</td>
</tr>
<tr>
<td>Number of unique 4-grams</td>
<td>55 (3.56%)</td>
</tr>
<tr>
<td>Number of unique 5+6-grams</td>
<td>9 (0.58%)</td>
</tr>
</tbody>
</table>

The statistics of the dataset

The inter-annotator proportion agreement results (week by week)
FACE: Feature-based Keyphrase Extraction

▷ Recast as a binary classification problem for a list of extracted candidates
▷ Candidates extracted based on POS patterns
▷ Trained a Logistic Regression model with the feature list:
  • **Linguistic**: POS (features 1-5), context (features 6-17), length of candidate
  • **Statistical**: frequency, collection frequency, tf-idf, language model
  • **External resources**: Wikipedia titles, ACM Computer Science keyphrase repository
  • **Section titles**

Non-binary numerical features are binned and discretized and represented as one-hot encodings

<table>
<thead>
<tr>
<th>System</th>
<th>AUC</th>
<th>Micro p</th>
<th>Micro r</th>
<th>Micro F1</th>
<th>Macro p</th>
<th>Macro r</th>
<th>Macro F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1: Random (*)</td>
<td>0.50</td>
<td>0.20</td>
<td>0.67</td>
<td>0.31</td>
<td>0.14</td>
<td>0.50</td>
<td>0.21</td>
</tr>
<tr>
<td>Baseline 2: Linguistics (*)</td>
<td>0.90</td>
<td>0.66</td>
<td>0.63</td>
<td>0.65</td>
<td>0.47</td>
<td>0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>Baseline 3: Statistics (*)</td>
<td>0.80</td>
<td>0.55</td>
<td>0.58</td>
<td>0.56</td>
<td>0.50</td>
<td>0.25</td>
<td>0.34</td>
</tr>
<tr>
<td>Baseline 4: External resources (*)</td>
<td>0.72</td>
<td>0.36</td>
<td>0.47</td>
<td>0.41</td>
<td>0.29</td>
<td>0.44</td>
<td>0.35</td>
</tr>
<tr>
<td>Baseline 5: Titles (*)</td>
<td>0.67</td>
<td>0.55</td>
<td>0.42</td>
<td>0.47</td>
<td>0.43</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td>Baseline 6: TextRank</td>
<td>–</td>
<td>0.17</td>
<td>0.10</td>
<td>0.17</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Baseline 7: TopicRank</td>
<td>–</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.11</td>
<td>0.28</td>
<td>0.16</td>
</tr>
<tr>
<td>Baseline 8: RAKE</td>
<td>–</td>
<td>0.15</td>
<td>0.13</td>
<td>0.14</td>
<td>0.07</td>
<td>0.63</td>
<td>0.12</td>
</tr>
<tr>
<td>Baseline 10: CopyRNN</td>
<td>–</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.26</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>Baseline 11: Humans (AMT)</td>
<td>–</td>
<td>0.40</td>
<td>0.38</td>
<td>0.39</td>
<td>0.29</td>
<td>0.55</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>FACE</strong></td>
<td>0.94</td>
<td>0.75</td>
<td>0.77</td>
<td>0.76</td>
<td>0.61</td>
<td>0.58</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Takeaway messages:
• Book’s rich structure provides useful features to identify keyphrases
• FACE framework can be applied for other domains
Limitation of Keyphrase Extraction

Muslim Women in Hijab Break Barriers: ‘Take the Good With the Bad’

When Ginella Massa, a Toronto-based TV reporter, recently accepted a request to host an evening newscast, she was not planning or expecting to make history for wearing a hijab. She was just covering for a colleague who wanted to go to a hockey game. And that’s how Ms. Massa, who works at CityNews in Toronto, became the first Canadian woman to host a newscast from a large media company while wearing the head scarf. [...] This new trend of inclusion occurs amid a more sinister one, as reported hate crimes against Muslims are on the rise in the United States and Canada. The F.B.I. says that a surge in hate crimes against Muslims has led to an overall increase in hate crimes in the United States; Muslims have borne the brunt of the increase with 257 recorded attacks. [...] In Canada, where Ms. Massa has lived since she was a year old, the number of reported hate crimes has dropped slightly overall, but the number of recorded attacks against Muslims has grown: 99 attacks were reported in 2014, according to an analysis by the news site Global News of data from Statistics Canada, a government agency. [...] 

Keywords: US; Islam; Fashion; Muslim Veiling; Women and Girls; (News media, journalism); Hate crime; Canada
Muslim Women in Hijab Break Barriers: ‘Take the Good With the Bad’

When Ginella Massa, a Toronto-based TV reporter, recently accepted a request to host an evening newscast, she was not planning or expecting to make history for wearing a hijab. She was just covering for a colleague who wanted to go to a hockey game. And that’s how Ms. Massa, who works at CityNews in Toronto, became the first Canadian woman to host a newscast from a large media company while wearing the head scarf. [...] This new trend of inclusion occurs amid a more sinister one, as reported hate crimes against Muslims are on the rise in the United States and Canada. The F.B.I. says that a surge in hate crimes against Muslims has led to an overall increase in hate crimes in the United States; Muslims have borne the brunt of the increase with 257 recorded attacks. [...] In Canada, where Ms. Massa has lived since she was a year old, the number of reported hate crimes has dropped slightly overall, but the number of recorded attacks against Muslims has grown: 99 attacks were reported in 2014, according to an analysis by the news site Global News of data from Statistics Canada, a government agency. [...] 

**keywords:** US; Islam; Fashion; Muslim Veiling; Women and Girls; (News media, journalism; Hate crime; Canada

---

Not All Keyphrases Are Extractable

- A non-negligible proportion of keyphrases are not present
- Annotators assign keyphrases by their relevance/importance, not presence

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Train</th>
<th>#Valid</th>
<th>#Test</th>
<th>Mean</th>
<th>Var</th>
<th>%Pre</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP20k</td>
<td>$\approx$514k</td>
<td>$\approx$20k</td>
<td>$\approx$20k</td>
<td>5.3</td>
<td>14.2</td>
<td>63.3%</td>
</tr>
<tr>
<td>INSPEC</td>
<td>–</td>
<td>1500</td>
<td>500</td>
<td>9.6</td>
<td>22.4</td>
<td>78.5%</td>
</tr>
<tr>
<td>KRAPIVIN</td>
<td>–</td>
<td>1844</td>
<td>460</td>
<td>5.2</td>
<td>6.6</td>
<td>56.2%</td>
</tr>
<tr>
<td>NUS</td>
<td>–</td>
<td>-</td>
<td>211</td>
<td>11.5</td>
<td>64.6</td>
<td>51.3%</td>
</tr>
<tr>
<td>SEMEVAL</td>
<td>–</td>
<td>144</td>
<td>100</td>
<td>15.7</td>
<td>15.1</td>
<td>44.5%</td>
</tr>
<tr>
<td>STACKEx</td>
<td>$\approx$298k</td>
<td>$\approx$16k</td>
<td>$\approx$16k</td>
<td>2.7</td>
<td>1.4</td>
<td>57.5%</td>
</tr>
</tbody>
</table>

(Yuan et al. 2018). One Size Does Not Fit All: Generating and Evaluating Variable Number of Keyphrases. ACL.
Method 2: Neural Keyphrase Generation

- Predicting keyphrases as language generation
  - Each keyphrase is actually a short sequence of tokens
  - We can train neural networks to learn to generate phrases in a data-driven way

**Input:** a **SEQ**uence of source text
**Output:** multiple **SEQ**uences of tokens, each sequence is a keyphrase

**Seq2Seq Learning!**

One source sequence

Multiple target sequences
Keyphrase Generation Models

- Seq2Seq + Copy Attention
  - Generate target keyphrase both abstractively and extractively

\[
P(w) = p_{abs} \cdot P_{abs}(w_{vocab}) + (1 - p_{abs}) \cdot P_{ext}(w_{src})
\]

(Meng et al. 2017). Deep Keyphrase Generation. ACL.
Keyphrase Generation (KPG)

Three types of training paradigms

- **One2One**: Output one single phrase at a time
- **One2Seq**: Output a sequence of multiple phrases at a time
- **One2Set**: Output a set of multiple phrases at a time

Rui Meng, Debanjan Mahata, Florian Boudin “From Fundamentals to Recent Advances: A Tutorial on Keyphrasification”, a half-day tutorial at the 44th European Conference on Information Retrieval (ECIR 2022) [https://keyphrasification.github.io/](https://keyphrasification.github.io/)

(Meng et al. 2017). Deep Keyphrase Generation. ACL.
(Yuan et al. 2018). One Size Does Not Fit All: Generating and Evaluating Variable Number of Keyphrases. ACL.
(Ye and Wang, 2018). Semi-Supervised Learning for Neural Keyphrase Generation. EMNLP.
(Meng et al. 2021). An Empirical Study on Neural Keyphrase Generation. NAACL.
(Ye et al. 2021) "One2Set: Generating Diverse Keyphrases as a Set. ACL."
KPG-One2One Vs. KPG-One2Seq

- Both are based on Sequence-to-Sequence Learning

[Source Sequence]=title+abstract
Language-specific Models in Multilingual Topic Tracking. Topic tracking is complicated when the stories in the stream occur in multiple languages. Typically, researchers have trained only English topic models because the training stories have been provided in English. In tracking, non-English test stories are then machine translated into English to compare them with the topic models. ...

[Target Sequence]=a list of keyphrases
[classification, crosslingual, Arabic, TDT, topic tracking, multilingual]
KPG-One2One

Data preparation - each data example is split to multiple text-keyphrase pairs
Source text is duplicated K times
Each pair contains only one keyphrase
Great waste in training, e.g. in KP20k 510K->2.78M

Original Data Point (k target phrases)

[Source]
Language-specific Models in Multilingual Topic Tracking. Topic tracking is complicated when the stories in the stream occur in multiple languages. Typically, researchers have trained only English topic models because the training stories have been provided in English. In tracking, non-English test stories are then machine translated into English to compare them with the topic models. …

[Target]
[classification, crosslingual, Arabic, TDT, topic tracking, multilingual]

Src-Tgt Pair for Training (k pairs)

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> classification </s>

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> crosslingual </s>

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> arabic </s>

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> TDT </s>

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> topic tracking </s>

[Source] Language-specific Models in Multilingual Topic Tracking,…
[Target] <s> multilingual </s>

(Meng et al. 2017). Deep Keyphrase Generation. ACL.
KPG-One2Seq

▷ Can the model generate multiple phrases directly?
  • Output
  • Let model to handle the interaction between phrases and avoid redundancy in output

▷ KPGen-One2Seq
  • Given ONE source text, the goal is to generate one SEQuence of concatenated keyphrases
  • Concatenate multiple target phrases as a sequence
  • The order of concatenation can be effective in performance

[Source Sequence]
Language-specific Models in Multilingual Topic Tracking.
Topic tracking is complicated when the stories in the stream occur in multiple languages. Typically, researchers have trained only English topic models because the training stories have been provided in English. In tracking, non-English test stories are then machine translated into English to compare them with the topic models. …

[Target Sequence]
classification, crosslingual, Arabic, TDT, topic tracking, multilingual

(Yuan et al. 2018). One Size Does Not Fit All: Generating and Evaluating Variable Number of Keyphrases. ACL.
(Ye and Wang, 2018). Semi-Supervised Learning for Neural Keyphrase Generation. EMNLP.
KPG-One2Seq

- One2Seq: But the order of concatenation may matter ...
- Several order options

[Source Sequence]=title+abstract
Language-specific Models in Multilingual Topic Tracking.
Topic tracking is complicated when the stories in the stream occur in multiple languages. Typically, researchers have trained only English topic models because the training stories have been provided in English. In tracking, non-English test stories are then machine translated into English to compare them with the topic models. ...

[Target Sequence]=keyphrases
[classification, crosslingual, Arabic, TDT, topic tracking, multilingual]

[Present Phrases] topic tracking, multilingual
[Absent Phrases] classification, crosslingual, Arabic, TDT
Datasets

- **Datasets**
  - KP20k (514k CS papers)
  - Inspec: 2,000 paper abstracts.
  - Krapivin: 2,304 papers with full-text and author-assigned keyphrases.
  - NUS: 211 papers with author-assigned and reader-assigned keyphrases
  - SemEval-2010: 288 articles from the ACM Digital Library

- **Relations**
  - Same domain, similar distribution
    - KP20k, Krapivin
  - Same domain, different distribution/annotation
    - Inspec, NUS, SemEval
  - Different domain
    - DUC (news article)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Train</th>
<th>#Valid</th>
<th>#Test</th>
<th>Mean</th>
<th>Var</th>
<th>%Pre</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP20k</td>
<td>≈514k</td>
<td>≈20k</td>
<td>≈20k</td>
<td>5.3</td>
<td>14.2</td>
<td>63.3%</td>
</tr>
<tr>
<td>MAGKP</td>
<td>≈2.7M</td>
<td>-</td>
<td>-</td>
<td>12.9</td>
<td>?</td>
<td>?%</td>
</tr>
<tr>
<td>Inspec</td>
<td>-</td>
<td>1500</td>
<td>500</td>
<td>9.6</td>
<td>22.4</td>
<td>78.5%</td>
</tr>
<tr>
<td>Krapivin</td>
<td>-</td>
<td>1844</td>
<td>460</td>
<td>5.2</td>
<td>6.6</td>
<td>56.2%</td>
</tr>
<tr>
<td>NUS</td>
<td>-</td>
<td>-</td>
<td>211</td>
<td>11.5</td>
<td>64.6</td>
<td>51.3%</td>
</tr>
<tr>
<td>SemEval</td>
<td>-</td>
<td>144</td>
<td>100</td>
<td>15.7</td>
<td>15.1</td>
<td>44.5%</td>
</tr>
<tr>
<td>StackEx</td>
<td>≈298k</td>
<td>≈16k</td>
<td>≈16k</td>
<td>2.7</td>
<td>1.4</td>
<td>57.5%</td>
</tr>
<tr>
<td>DUC</td>
<td>-</td>
<td>-</td>
<td>308</td>
<td>8.1</td>
<td>?</td>
<td>97.5%</td>
</tr>
</tbody>
</table>

Table 1: Statistics of various datasets. Mean and Var indicate the mean and variance of target phrase numbers, %Pre denotes percentage of present keyphrases.
Learning Curve (F@10) of In-Domain Datasets

Impressions

▷ One2one converges faster than one2seq, and performs better.

▷ Valid curve is in line with test curve, and always better.
A Closer Look at All Datasets - Present

Impressions

- One2One performs better on in-domain datasets (KP20K and Krapivin) and NUS.
- One2Seq’s transferability looks better: on Inspec and DUC (news) it outperforms One2One significantly.
One2One performs much better than One2Seq due to its superior ability in generating unique phrases.

**Impressions**

**Takeaway messages:**
- Keyphrase generation is a more powerful task modeling
- But its effective generation methods are still opening questions
Outline

- Basics of Keyphrases: Definitions and Importance
- Identification of Keyphrases: Extraction and Generation
- Applications of Keyphrases: knowledge unit for supporting student learning
- Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
- Applications of Keyphrases: knowledge unit for interactive machine learning
- Conclusions

Recommending Remedial Readings Using Student Knowledge State
Khushboo Thaker, Lei Zhang, Daqing He, Peter Brusilovsky
Intelligent Textbooks

Sections in chapter

Chapter / Module

Evaluation
Self-Reflection

Assessment

A, B, C, D

Learning Curve Height

Background
Intelligent Textbooks

Sections in chapter

Remedial Content

Failure

Assessment

Evaluation

Self-Reflection

Chapter/Module

Objective
Remedial Recommendation

Challenge

sections

Computer Assisted Instructions (CAI)

Mapping

Questions
Simple Text-based Similarity and its challenge

Challenges:
- **Advance Content** - Student lack pre-requisites
- **Redundant Content** - Student already mastered
- **Not Personalized** to students need

Example Text: *Information retrieval* is the activity of obtaining *information system* resources that are relevant to an *information need* from a collection of those resources. *Searches* can be based on full-text or other *content-based indexing*. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for the *metadata* that describes data, and for databases of texts, images or sounds.
Motivation

Student knowledge level

Extracted Keyphrases

Example Text:
Information retrieval is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds.
Research Questions:

1. Does the **concept-based representation** of educational content help perform remedial recommendation, either by acting alone or in combination with the content-based recommendation?

2. Does the augmentation of **student knowledge** on concept-based representation help in provide personalized remedial recommendations?

3. Can we use **automated key phrase extraction** techniques to generate concept-based representation?
Each Section is annotated with identified keyphrases as domain concepts.
ReadingCircle: Information Retrieval Textbook

Each Section is annotated with identified keyphrases as domain concepts.
Remedial Recommendation

Sections in chapter

Chapter/Module
Concept-based Remedial Recommendation

- For gold standard – we used experts to match mapped each question to relevant section.
- Dataset available on datashop¹

¹https://pslcdatashop.web.cmu.edu/Project?id=637
Knowledge based Remedial Recommendation

Student knowledge

Student modeling

output

PFA with Concepts as Knowledge Components

Evaluation

Self-Reflection

Chapter / Module

PFA – Performance Factor Analysis
Knowledge based Remedial Recommendation

Student modeling → Remedial Content → Failure

Assessment

Chapter / Module

Evaluation
Self-Reflection

PFA – Performance Factor Analysis

PFA with Concepts as Knowledge Components
Knowledge-based Remedial Recommendation

Student dynamic knowledge state

Question

Book Sections

Cosine Similarity

Ranking

Section 1
Section 2
Section 3

1 https://pslcdatashop.web.cmu.edu/Project?id=637
## Results: Keyphrases as Concepts

<table>
<thead>
<tr>
<th>Method</th>
<th>Text-based</th>
<th>Concept-based</th>
<th>Knowledge-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAP@5</td>
<td>MAP@5</td>
</tr>
<tr>
<td>Text Similarity</td>
<td>0.74</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Expert Concepts</td>
<td>-</td>
<td>0.8618</td>
<td>0.8390</td>
</tr>
<tr>
<td>TextRank</td>
<td>-</td>
<td>0.8314</td>
<td>0.8397</td>
</tr>
<tr>
<td>CopyRNN</td>
<td>-</td>
<td>0.8466</td>
<td>0.8405</td>
</tr>
<tr>
<td>TopicRank</td>
<td></td>
<td>0.8990*</td>
<td>0.8885*</td>
</tr>
</tbody>
</table>

All are better than text-based

Bold indicates significantly better than baseline with expert generated concepts
Knowledge-based Remedial Recommendation
It should recommend different thing to different people?

- For 86% of quizzes students are recommended different sections
- There are cases where a failure on a quiz is mapped to more than 15 unique sections
- This shows that student knowledge does provide personalized recommendations

**Takeaway messages:**
- Students’ learning can be helped with knowledge units identified from text
- Keyphrases can act as useful knowledge units in supporting students’ learning
Outline

• Basics of Keyphrases: Definitions and Importance
• Identification of Keyphrases: Extraction and Generation
• Applications of Keyphrases: knowledge unit for supporting student learning
• Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
• Applications of Keyphrases: knowledge unit for interactive machine learning
• Conclusions

Weakly Supervised Medical Entity Extraction and Linking for Chief Complaints
Zhimeng Luo, Zhendong Wang, Rui Meng, Diyang Xue, Adam Frisch and Daqing He
Chief Complaints

The beginning of physician’s diagnosis process in emergency department (ED) is guided by the patient’s chief complaint (CC).

Chief Complaint is a record to summarize:

- reason for encounter
- current symptoms
- medical history

Chief Complaint: migraine with neck/back pain, fever

Patient → Nurse → Doctor

- Triage
  - Summarize CC
  - Assign priority level
- Order tests
- Diagnosis
- Treatment
Chief Complaints

Chief complaint record/instance: ha light headed fatigue r arm pain

Chief complaint entity mentions: (Span: location of each mention)

Chief complaint entity concepts: (from HaPPy’s ontology)

headache ; dizziness ; fatigue ; arm pain

Characteristics of chief complaint:

- Short free-text descriptions, with large variation (abbreviations, synonyms, ...)
- A record may contain one or multiple CC entity mentions
- Span of each concept is important
  - Doctors want to know span along with the concepts
  - To summarize entity variants to improve the existing ontology
HaPPy ontology

- The first publicly available large-scale CC ontology
- Containing 692 unique concepts, 2,118 synonyms, and 30,613 descriptions.

- We found that
  - Direct match result in low matching recall following HaPPy’s instruction across health care system (on UPMC corpus).
  - automatic CC extraction and linking is necessary
  - use it as ground truth concept labels in our project
Limitations of Related Work

Limitations:
- Cannot identify multiple CC concepts within a record
- Cannot identify the span of each CC entity mentions
- Lack of large-scale annotations containing span information

We propose to view the task as:

- **entity extraction**
  - identifying the actual mention span of each entity in a free text
- **entity linking**
  - linking each entity mention to a concept in a CC ontology
- **Weak supervision**
  - Identifying noisy patterns in text without manual annotations?
Proposed method: WeSEEL

Figure 1: Overview of our proposed method WeSEEL (Weakly Supervised Entity Extraction and Linking in chief complaint).
Weak label generation

To assign concept labels with corresponding span information for model training, a **Split-and-Match** algorithm is proposed as follows:

- **Exact** string matching (HaPPy ontology)
- **Approximate** string matching (QuickUMLS)
  - resolve misspelling and lexical variations
- **Embedding**-based matching (fastText)
  - enables semantic matching

(a) Process flow of weak label generation. Three examples are shown and successfully matched to concepts in the ontology at different stages (indicated in green box).
Two-step model

- Entity Mention **Extraction** (extraction model)
- **Linking** Entities to Ontology (linking model)
Formulate as a **sequence labeling** problem that follows the BIO tagging:

- e.g., "10 wks/n/v/d" -> "10 wks / n / v / d"

\[
B \ I \ O \ B \ O \ B \ O \ B
\]

- BERT token classification model (CCME)
- Soft label
  - adjust the label smoothing to accommodate the weak span labels
  - For each word in a chunk, set the probability of a weak target label as the **similarity** between a chunk and its corresponding ontology concept

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Adjusted Token Weight</th>
<th>Adjusted Target Vector (B/I/O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chest pain and</td>
<td>0.9</td>
<td>[0.9, 0.0, 0.1]</td>
</tr>
<tr>
<td>chest pain and</td>
<td>0.9</td>
<td>[0.0, 0.9, 0.1]</td>
</tr>
<tr>
<td>chest pain and</td>
<td>0.0</td>
<td>[0.0, 0.0, 1.0]</td>
</tr>
</tbody>
</table>
Linking Entities to Ontology

• Link each entity mention extracted from CC records to a given ontology through a classification model.

• Use BiLSTM as basic layer.

• Propose two additional input embeddings along with the major mention word embedding:
  • Surrounding context embedding
    • to consider context information
  • Character embedding of mention tokens
    • to consider lexicon variations
Data set

• 1,232,899 free-text CC records were collected from UPMC Health Service system
• covering the period of 2015 to 2017 from 15 hospitals.
• All EDs use the same electronic health record system, but do not mandate a specific data entry format.
• A test set of 1,013 random instances was annotated by an ED clinician.
• HaPPy ontology was selected as the target ontology to link entities
  • shrunk the original ontology (692 to 501 concepts) by removing child nodes that have no significant clinical difference with their parent node
  • e.g., “ruq abdominal pain” (“ruq” means right upper quadrant) were folded to parent node “abd pain”
• All data have been checked by IRB
## Data set

<table>
<thead>
<tr>
<th>No. cc</th>
<th>instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>1</td>
<td>388</td>
</tr>
<tr>
<td>2</td>
<td>371</td>
</tr>
<tr>
<td>3</td>
<td>143</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

No. concepts in each record in test set.

<table>
<thead>
<tr>
<th>Punctuation</th>
<th>#Tokens</th>
<th>Conditions</th>
<th>#Instances</th>
<th>Percentage</th>
<th>#Sampled</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/ Single</td>
<td>7,739</td>
<td>Only contains &quot;/&quot;</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Single</td>
<td>1,285</td>
<td>Only contains &quot;,,&quot;</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Single</td>
<td>529</td>
<td>Only contains &quot;,\br&quot;</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Single</td>
<td>459</td>
<td>Only contains &quot;,...&quot;</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Single</td>
<td>1,097</td>
<td>Others</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Single</td>
<td>11,109</td>
<td>-</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>154,457</td>
<td>Only contains &quot;,,&quot;</td>
<td>324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>78,129</td>
<td>Only contains &quot;/&quot;</td>
<td>164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>4,017</td>
<td>Only contains &quot;,...&quot;</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>3,248</td>
<td>Only contains &quot;,\br&quot;</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>2,154</td>
<td>Only contains &quot;,+&quot;</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>1,902</td>
<td>Only contains &quot;,,&quot;</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>1,895</td>
<td>Only contains &quot;,.,&quot;</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>1,695</td>
<td>Only contains &quot;,&quot;</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>1,599</td>
<td>Only contains &quot;,&amp;&quot;</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>17,555</td>
<td>Contains &quot;,,&quot; and &quot;,/&quot;</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>11,182</td>
<td>Others</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/ Multiple</td>
<td>277,833</td>
<td>-</td>
<td>583</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o Single</td>
<td>5,108</td>
<td>Others</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o Multiple</td>
<td>19,078</td>
<td>Only contains &quot;and&quot;</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o Multiple</td>
<td>1,292</td>
<td>Only contains &quot;at&quot;</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o Multiple</td>
<td>168,118</td>
<td>Others</td>
<td>353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o -</td>
<td>193,596</td>
<td>-</td>
<td>407</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Distribution of unique instances with conditions. The test set is sampled based on the distribution of conditions, shown as #Sampled. The percentage less than 2% are not shown in the table.
Evaluation metrics

• We adopt the evaluation protocol of SemEval 2013 task 9.1
• Mention extraction
  “Partial” mode (partial boundary match, regardless of the type)
  “Exact” mode (exact boundary match, regardless of the type)
• Entity linking
  “Entity type” mode (partial boundary match and correct entity type)
Test results of entity extraction

<table>
<thead>
<tr>
<th>Models</th>
<th>Partial Match</th>
<th>Exact Match</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>S&amp;M (HaPPy)</td>
<td>95.81</td>
<td>36.90</td>
</tr>
<tr>
<td>S&amp;M (QuickUMLs)</td>
<td>81.15</td>
<td>46.04</td>
</tr>
<tr>
<td>S&amp;M (Embedding)</td>
<td>69.64</td>
<td><strong>57.36</strong></td>
</tr>
<tr>
<td>CCME-LSTM</td>
<td>78.45</td>
<td>48.26</td>
</tr>
<tr>
<td>CCME-BERT</td>
<td>81.37</td>
<td>53.43</td>
</tr>
<tr>
<td>CCME-BERT (soft)</td>
<td>83.41</td>
<td>56.70</td>
</tr>
<tr>
<td>CCME-ClinicalBERT (soft)</td>
<td>83.35</td>
<td>56.46</td>
</tr>
<tr>
<td>CCME-BERT (soft) + S&amp;M (HaPPy)</td>
<td><strong>96.28</strong></td>
<td>44.86</td>
</tr>
<tr>
<td>CCME-BERT (soft) + S&amp;M (QuickUMLs)</td>
<td>86.13</td>
<td>51.82</td>
</tr>
</tbody>
</table>

Table 4: Entity extraction performance of different models. Scores are computed in Partial Match and Exact Match mode of SemEval’13. The best/2nd-best scores in each column are in bold/underlined.

- Most neural models outperform the matching-based methods, indicating that machine learning models can learn task-relevant inductive bias from weak labels.
Test results of entity linking

<table>
<thead>
<tr>
<th>#</th>
<th>Extraction Model</th>
<th>Linking Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.1</td>
<td>†S&amp;M (S.1)</td>
<td>†S&amp;M (S.1)</td>
<td>98.02</td>
<td>37.75</td>
<td>54.51</td>
</tr>
<tr>
<td>s.2</td>
<td>†S&amp;M (S.1 + S.2)</td>
<td>†S&amp;M (S.1 + S.2)</td>
<td>79.34</td>
<td>45.02</td>
<td>57.44</td>
</tr>
<tr>
<td>s.3</td>
<td>†S&amp;M (S.1 + S.2 + S.3)</td>
<td>†S&amp;M (S.1 + S.2 + S.3)</td>
<td>57.73</td>
<td>47.54</td>
<td>52.44</td>
</tr>
<tr>
<td>n.1</td>
<td>CCME-BERT (soft)</td>
<td>fastText (single-label)</td>
<td>92.93</td>
<td>21.11</td>
<td>34.41</td>
</tr>
<tr>
<td>n.2</td>
<td>CCME-BERT (soft)</td>
<td>fastText</td>
<td>89.27</td>
<td>26.69</td>
<td>41.09</td>
</tr>
<tr>
<td>n.3</td>
<td>CCME-BERT (soft)</td>
<td>BERT</td>
<td>83.48</td>
<td>34.52</td>
<td>48.84</td>
</tr>
<tr>
<td>n.4</td>
<td>CCME-BERT (soft)</td>
<td>CCEL</td>
<td>84.36</td>
<td>45.22</td>
<td>58.88</td>
</tr>
<tr>
<td>b.1</td>
<td>QuickUMLS</td>
<td>MedType (EHR)</td>
<td>53.32</td>
<td>23.09</td>
<td>32.23</td>
</tr>
<tr>
<td>m.1</td>
<td>†S&amp;M (S.1 + S.2)</td>
<td>fastText (single-label)</td>
<td>89.07</td>
<td>15.77</td>
<td>26.79</td>
</tr>
<tr>
<td>m.2</td>
<td>†S&amp;M (S.1 + S.2)</td>
<td>fastText</td>
<td>86.49</td>
<td>19.52</td>
<td>31.85</td>
</tr>
<tr>
<td>m.3</td>
<td>CCME-BERT (soft)</td>
<td>†S&amp;M (S.1)</td>
<td>97.86</td>
<td>45.60</td>
<td>62.20</td>
</tr>
<tr>
<td>m.4</td>
<td>CCME-BERT (soft)</td>
<td>†S&amp;M (S.1 + S.2)</td>
<td>85.64</td>
<td>51.53</td>
<td>64.34</td>
</tr>
<tr>
<td>m.5</td>
<td>CCME-BERT (soft)</td>
<td>†S&amp;M (S.1 + S.2) + CCEL</td>
<td>86.28</td>
<td>55.43</td>
<td>67.49</td>
</tr>
</tbody>
</table>

Table 5: Entity linking performance. Scores are computed in Entity Type mode of SemEval'13. The best/2nd-best scores in each column are in bold/underlined. †S.1, S.2, S.3 refer to string matching algorithms in Figure 1(a).

- CCME-BERT models are good at identifying entity mentions, while matching methods are good at classifying concepts given identified mentions
Effect of Weak Supervision

We simulate a fully supervised setting: 80% training, 20% testing from annotated test set (report average scores from 5-fold CV)

- Supervised: training models with annotated data only;
- Fine-tuning: pre-train using weak labels and fine-tune it with the annotated data.

<table>
<thead>
<tr>
<th>Training</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>WeakSup</td>
<td>83.41</td>
<td>56.70</td>
<td>67.51</td>
</tr>
<tr>
<td>Supervised</td>
<td>77.76</td>
<td>89.66</td>
<td>83.29</td>
</tr>
<tr>
<td>Fine-tuning</td>
<td>82.25</td>
<td>85.98</td>
<td>84.07</td>
</tr>
</tbody>
</table>

Table 7: Extraction performance (Partial Match) of CCME-BERT (soft) with three training strategies.

- Trained with little annotated data, CCME-BERT achieve decent results on mention extraction
- Pre-training the model with weak labels can be beneficial
EHR: admission note

- EHR Notes contains narrative information about a patient’s current and past medical history.
  - Many types of notes in the EHR including: admission notes, assessments, SOAP notes, exams, reports, and etc.
- **Admission notes** document the reasons why a patient is being admitted to a hospital or other facility, the patient's baseline status, and the initial instructions for that patient's care. Its important components are
  - **Chief Complaints (CC):** "abdominal pain"
  - **History of Present Illness (HPI):** "Pt is a 30 yo female (with a PMH of x and y) presenting with a 3 hour history of abdominal pain..."
  - **Review of Systems (ROS):** “immunologic : negative. \n musculoskeletal : right lower extremity pain and swelling.. ”
  - **Medical Decision Making (assessment):** similar to the first line of the HPI, but with a greater emphasis on clinical reasoning.
  - **Diagnosis and Plan**
Examples note 1

● **CC section**
  ○ “easy bruising (ecchymosis), rle pain (leg pain) and swelling (leg swelling)”

● **ROS section**
  ○ “musculoskeletal: right lower extremity pain (leg pain) and swelling (leg swelling)”
  ○ “ integumentary: petechiae”

● **HPI section**
  ○ “patient is a 39 y/o female with pmh + lupus anticoagulant c/b recurrent pes now on warfarin, kidney stones, asthma (asthma exacerbation), seizure disorder, bipolar disorder, hypothyroidism who presents for evaluation of right lower extremity pain (leg pain) and easy bruising (ecchymosis). patient reports recent fall on saturday 5/6/17, tells me that her shoes "caught up" and then she tripped and fell to the ground hitting her back and her right lower leg. denies trauma to the head or loss of consciousness…”

● **Medical Decision Making section**
  ○ “patient presents for evaluation of increasing ecchymosis, right leg pain and swelling. i suspect this is primarily related to significant hematoma in the right leg though there is concern for dvt (prior to knowledge of her inr). right lobe chevy dopplers obtained shows no evidence of dvt. her inr was found to be significantly supratherapeutic...because of pain control, need to trend hemoglobin given the possibility of significant bleeding into the right leg.”

● **Diagnosis**
  ○ “supratherapeutic inr, right lower extremity hematoma, right leg swelling
Building knowledge graph
Showcase the usage of the graph

Neighbor Extraction

Diagnosis Prediction

Takeaway messages:
- Chief complaints are noisy yet important keyphrases in clinical text
- Our method paves a foundation for further exploration of clinical text using CC as knowledge units

Outline

- Basics of Keyphrases: Definitions and Importance
- Identification of Keyphrases: Extraction and Generation
- Applications of Keyphrases: knowledge unit for supporting student learning
- Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
- Applications of Keyphrases: knowledge unit for interactive machine learning
- Conclusions

Characterizing Dementia Caregivers’ Information Exchange on Social Media: Exploring an Expert-Machine Co-Development Process Zhendong Wang¹, Ning Zou¹, Bo Xie², Zhimeng Luo¹, Daqing He¹, Robin C. Hilsabeck², Alyssa Aguirre²
Background

- Alzheimer’s disease and related dementias (ADRD) are a major public health concern
- In the U.S., about 5.6 million Americans age 65 and over were living with ADRD in 2019[1]
- Caregiving for people with ADRD is stressful[2][3][4]
Background

• Social media platforms have introduced novel mechanisms supporting online health information seeking and sharing.

• Research on ADRD caregivers’ information exchange via social media platforms remains limited.
Previous works

• **Expert Analysis**: relies on human experts to manually analyze social media content
  - Accurate but time consuming, costly, and problematic for large amounts of rapidly growing social media data.
  - Social media is also new for human experts
• **Automatic Exploration**: uses machine learning or text mining algorithms
  - Able to overcome these limitations of expert analysis
  - Requires large annotated data from human experts
  - Lacks iterative interaction or knowledge exchange between human experts and automatic algorithms
• Interactive Machine Learning (IML):
  • IML utilizes learning loop with interaction from human experts to iteratively increase performance of machine learning model with less human efforts
  • IML has been applied successfully in wide range of domains\[5\][6][7] but not yet in health information exchange.
Proposed method: EMC Process

- **Expert-Machine Co-development (EMC) Process:**
  - Create a Health Information Want (HIW) framework to analyze the category and keywords of ADRD online posts
  - IML based interactive process with rich interactions
  - Maximize the strengths of both human experts and automatic algorithms
  - Minimize human efforts

- **Components:**
  - **Component 1:** Expert Analysis of ADRD Caregivers’ Information Exchange
  - **Component 2:** Automatic Exploration (AE) of ADRD Caregivers’ Information Behaviors

*Figure 1. EMC Process*
Research Aims

• Aim 1: What ADRD-related information do caregivers exchange on social media?

• Aim 2: How an interactive learning system can be designed to enable the EMC process

• Aim 2: What roles can keyphrases extracted from online posts can play to help both human experts and the machine learning system?
Component 1: Expert Analysis of ADRD Caregivers Information Exchange

• **Goal:** create an initial framework for ADRD information exchange analysis

• **Collecting Data:** 823 posts from reddit group of Alzheimer

• **Health Information Wants framework (HIW-ADRD) development:**
  • 7 Categories
  • 176 keyphrases
  • 200 manually annotated posts

<table>
<thead>
<tr>
<th>Type of information</th>
<th>Sample keywords</th>
<th>Number(%) of posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment / Medication / Prevention</td>
<td>Drug; oriental; acupuncture; vitamins</td>
<td>8(4)</td>
</tr>
<tr>
<td>Characteristics of / Experience with the health condition / Diagnostic procedures</td>
<td>Diagnosis; complication; cause; prognosis; process; symptom; memory loss; lab test; MRI; PET; blood</td>
<td>17(8.5)</td>
</tr>
<tr>
<td>Daily care for a patient at home / Care for a caregiver (practical strategies or tips, not psychosocial)</td>
<td>Wandering; bath; hygiene; sleep; eat; driving</td>
<td>28(14)</td>
</tr>
<tr>
<td>Care transition and coordination / End-of-life care (practical, not psychosocial)</td>
<td>Adult day care; rehab; hospital; memory care; nursing home; hospice</td>
<td>13(6.5)</td>
</tr>
<tr>
<td>Psychosocial aspects</td>
<td>Stress; lonely; heartbreaking; overwhelmed; venting</td>
<td>66(33)</td>
</tr>
<tr>
<td>Resources / Advocacy / Scientific updates / Research participation</td>
<td>Lobby; fundraising; clinical trial; news; article; scientist</td>
<td>63(31.5)</td>
</tr>
<tr>
<td>Legal / Financial / Insurance</td>
<td>Power of attorney; POA; living will; Medicare; Medicaid</td>
<td>5(2.5)</td>
</tr>
</tbody>
</table>

**Table 1.** The HIW-ADRD 3.0 framework
Component 2: Automatic Exploration (AE) of ADRD Caregivers’ Information Behaviors

• Goal: Improve the HIW-ADRD framework
  • Tuning existing keyphrases
  • Help human experts discover better keyphrases
  • Improve the accuracy of model

• AE Process
  • Initial Model training
  • Interactive Learning loop
    • Keyphrases tuning recommendation
    • AE assisted exploration and feedback
    • Annotation and stop recommendations
Component 2.1: Initial Model training

- **Model**
  - **Input:** document representations of keyphrases and their frequency
  - **Output:** HIW-ADRD categories
- **Initial training dataset:** annotation from result of Component 1

---

**Figure 2. Automatic Exploration Process**

- Initial Model Training
- Keyphrases Tuning Recommendation
- AE assisted exploration and feedback
- Annotation Recommendation
- Stop?
  - No
  - Yes: End
Component 2.2: Keyphrases tuning recommendation

- recommendation criteria:
  - Mutual Information (MI) Score (normalized) between keyphrases and category
  - Importance(I) Score between keyphrases and model
  - Keyphrase Frequency (KF) for keyphrases in posts

- 4 Keyphrases Tuning (KT)
  - potentially good (PG): high MI, low I
  - potentially bad (PB): high I, low MI
  - low frequent (LF): too small KF
  - New Keywords (NK): Not in the existing framework, but has potentially high MI and enough KF

Figure 2. Automatic Exploration Process
Component 2.3: AE assisted exploration and feedback

Figure 2. Automatic Exploration Process

Figure 3. Interactive Auto Exploration Interface (IAEI)
Component 2.4: Annotation and stop recommendations

- **Annotation recommendation**
  - We rank the unannotated posts according to their **Aggregated MI(AMI) score**
  - Pick up top $n$ (10 in our case) to let annotator do more annotation

- **Stop recommendation**
  - Accuracy of ML not increase
  - normalized discounted cumulative gain (nDCG) between the keyword’s MI score ranking (descending) and the keyword I score ranking not increase.

\[
AMI(p) = \sum_{i=1}^{n} MI(t_i, c_t)
\]

**Figure 5.** Aggregated MI score formula
Experiment Results

• **Result**
  - IML improve the performance
  - KT doesn’t always improve the performance but with IML, it achieves best performance
  - Human experts annotated just 40 more posts

• **HIW-ADRD 3.1 framework**
  - discovered 7 better keyphrases to replace the existing one
  - removed 6 keyphrases
  - reviewed 25 new keyphrases and include 15 of them

<table>
<thead>
<tr>
<th>Training group</th>
<th>Dataset</th>
<th>Xgboost</th>
<th>SVM</th>
<th>Naïve Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>TRAIN</td>
<td>0.870</td>
<td>0.640</td>
<td>0.705</td>
</tr>
<tr>
<td></td>
<td>TEST</td>
<td>0.281</td>
<td>0.351</td>
<td>0.281</td>
</tr>
<tr>
<td>ID+RD (benchmark)</td>
<td>TRAIN</td>
<td>0.854</td>
<td>0.618</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td>TEST</td>
<td>0.333</td>
<td>0.404</td>
<td>0.333</td>
</tr>
<tr>
<td>ID + IML</td>
<td>TRAIN</td>
<td>0.887</td>
<td>0.637</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>TEST</td>
<td>0.509</td>
<td>0.421</td>
<td>0.386</td>
</tr>
<tr>
<td>ID+RD+KT (benchmark)</td>
<td>TRAIN</td>
<td>0.851</td>
<td>0.630</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>TEST</td>
<td>0.316</td>
<td>0.439</td>
<td>0.316</td>
</tr>
<tr>
<td>ID+IML+KT</td>
<td>TRAIN</td>
<td>0.892</td>
<td>0.658</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>TEST</td>
<td>0.544*</td>
<td>0.456</td>
<td>0.421</td>
</tr>
</tbody>
</table>

*Table 3. model accuracy*
Usability Interview with Human Experts

- **Strengths**
  - The visualization of keyphrase recommendations and navigation
  - The category distribution for search
  - The highlight keyphrases in posts

- **Weaknesses**
  - It takes time for domain experts to understand MI and I score
  - Some keyphrases recommendation is bad because of word normalization

**Takeaway messages:**
- Interactive ML can be used to enable expert machine collaboration
- Keyphrases can be the focus of the interactions to enable the collaboration
Outline

• Basics of Keyphrases: Definitions and Importance
• Identification of Keyphrases: Extraction and Generation
• Applications of Keyphrases: knowledge unit for supporting student learning
• Applications of Keyphrases: knowledge unit for recognizing patients’ concerns
• Applications of Keyphrases: knowledge unit for interactive machine learning
• Conclusions
Conclusions

▷ Keyphrases are short noun phrases to summarize and highlight important information in a piece of text
  • Different to and related to words and concepts
▷ Keyphrases can provide unique contributions as computable knowledge unit
  • Only a few possible applications are presented, many more can be explored
▷ Keyphrases can take in different roles in the text
  • Still open questions on how to identify and make use of their roles
Acknowledgement

Collaborators who made the mentioned work possible

▷ School of Computing and Information
  • Rui Meng, Khushboo Thaker, Zhimeng Luo, Hung Chau, Zhendong Wang, Ning Zou, Mengdi Wang, Peter Brusilovsky, Diyang Xue

▷ UPMC
  • Adam Frisch

▷ University of Texas Austin
  • Bo Xie, Robin C. Hilsabeck, Alyssa Aguirre

Funding Agencies that partially supported the work

• National Science Foundation, Amazon, National Institute of Health, University of Pittsburgh
Thank you & Questions