Can Generative AI aid or solve the Patent Summarization puzzle?
Invited speaker at 1st Workshop patent4Science
Outline

• Academic Overview
  • Text summarization algorithm
  • Patent text summarization

• What can Generative AI do with confident?

• How do we solve the puzzle to automatic generate patent summaries?
  • What is the information requirement of good patent summarization
  • How build a patent summarization algorithm that meet the need of the user and the different purpose
Text Summarization Algorithms
(~990 articles on summarization ACL)

Patent text Summarization

• There exists approximately 60 to 80 research articles related to patent summarization

• Used for query formulation
  • Evaluation the ability to retrieve relevant patent

• Used for to summarization
  • Evaluation abstract is defined as Gold Standard

• TRIZ
  • Manual or semi manual handcrafted data set
Query formulation

  - Discovered that the segment to use as query formulation is the Brief summary segment of US patent.
    - Evaluation on TREC-Chem

  - Evaluation on CLEF-IP Passage Patent retrieval
Automatic Query formulation
- From text segment to automatic Boolean query

Example of automatic query generation

```
<QUERY>
(conure OR clutch OR connectivity OR mnofs OR flp OR dlnr OR slippage OR anda OR rotational OR acceleration OR backlash OR subordinate OR estimating OR ure OR brake OR torque OR stopped OR vehicle OR wheel OR command OR outputting OR estimated OR book OR arc OR driving OR pedal OR wheels OR shaft OR prohibiting OR determining OR estimated OR prescribed OR stopping OR elapse OR output OR controller OR rotating OR accelerate OR AND

("vehicle driving force control apparatus" OR "dr OR "clutch connection command" OR "rear wheel path" OR "output rotational speed" OR "input to detected parameter" OR "generation load torque" OR "determination occurrence" OR "four-wheel drive proceed" OR "wheel speed sensor" OR "output s backlash elimination" OR "drive mode switch" OR range OR "transition time" OR "wheel speed" OR connection OR "motor torque" OR "generator k high rate" OR "electric motor" OR "throttle open force" OR "connected state" OR "previous equate prescribed rotational speed difference" OR "12 disconnected state OR "electric clutch" OR "to

</QUERY>
```

Automatic query expansion terms

- brake pedal:
- vehicle operating pedal,
- conventional hydraulic brake system
- pedal devices
- position brake actuating member
- brake actuating member
- hydraulically-assisted rack pinion steering gear
- brake operating member
- conventional braking system
- pedal pair

- accelerator pedal
- case pedal device
- pedal device

Claims (1)

1. The ornamental design for an automobile body, as shown and described.
Proof-of-concept
-Passage retrieval performance on CLEF-IP 2013 text collection.

<table>
<thead>
<tr>
<th>Method</th>
<th>PRES@100</th>
<th>Recall@100</th>
<th>MAP@100</th>
<th>Prec(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Albarede et. al. 2021)</td>
<td>0.461</td>
<td>0.603</td>
<td>0.173</td>
<td>0.231</td>
</tr>
<tr>
<td>(Andersson et. al., 2016)</td>
<td>0.444</td>
<td>0.560</td>
<td>0.187</td>
<td>0.282</td>
</tr>
<tr>
<td>(Andersson et. al., 2017)</td>
<td>0.558</td>
<td>0.647</td>
<td>0.269</td>
<td>0.207</td>
</tr>
<tr>
<td>(Luo J and Yang H. 2013)</td>
<td>0.433</td>
<td>0.540</td>
<td>0.191</td>
<td>0.213</td>
</tr>
</tbody>
</table>

Andersson L, Rekabsaz N., Hanbury A. (2017) Automatic query expansion for patent passage retrieval using paradigmatic and syntagmatic information The first WiNLP Workshop co-located with the Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver
Not one algorithm is suitable for all technical fields
Patent Summarization (topicalization)

  • Evaluation - manually written summaries (code and data set not accessible for research)
  • BIGPATANT


• Casola S, Lavelli A, (2022) Summarization, simplification, and generation: The case of patents, Expert Systems with Applications,
TRIZ
(Russian acronym corresponding to “Theory of Inventive Problem Solving”)

• DeWulf, S. "Patent data driven innovation logic." (2020). Imperial College London (UK)


• Chikkamath R., Ramsinh Parmar V. Hewel C. PaSA: A Dataset for Patent Sentiment Analysis to Highlight Patent Paragraphs (ICAIL 2022, SINGAPORE MAY 05-06, 2022)
What can Generative AI do with confident?
Generative AI models as chatGPT are as teenager – the behave as grown-ups but are still just in the beginning of their journey.

I compare it with asking teenager to clean the room. After 10min the teenager is back claiming it’s done, and you do inspection, and you find that everything is just swept under the bed.

You ask to re-do the cleaning-up and after 20 min the teenager is back with claim at least now it’s done. You do a second inspection, and it turns out everything in the wardrobe this time around.

So, you end up monitoring the teenager /AI generative model and give clear instruction and in every step, you verify that the task has been conducted according to instruction.

But sometimes it does happen occasionally that sometimes you ask the teenager/Generative AI model a task and it does it perfect the first time around!
Data is the king maker
Data

• Large Language Models (LLM) are not knowledge-based by themselves; they are intermediaries.
  • For specific text domains, these models can provide additional knowledge and contextual information, as well as offer natural language inference. However, verification of correctness is always required.

• Therefore...
  • Use your own knowledge based
    • Collect, categories and curate your data
    • LLM models can assist in the curation step, but verifying correctness is a necessity.
  • Always ensure to allocate time for data curation and normalization.
    • In a successful text mining application, 60% of the time is dedicated to collecting, normalizing, and data annotation, while only 40% is allocated to software development.
Designing is the Queen

- Control the flow of the data
  - Modelling the data
  - Amount of text
  - Annotation
    - Seed and Anti-seed
  - Feedback loops
Technology is just the pawns

- Utilize technologies that are suitable for the given task.
- Use the technologies that are applicable for the data
- Most likely you will arrive at an ensemble solution
Can LLM models assist in the curation step?

- Text augmentation
- Pre-Seed
  - Domain name entity
  - Text segment comparison
A turbine is a rotary mechanical device that extracts energy from a fluid flow and converts it into useful work," the following device concepts can be extracted:

**Extraction & explanation from chatGPT 3.5**

- **Turbine**: The primary device being described.
- **Rotary Mechanical Device**: Indicates that the turbine has a rotating mechanism.
- **Extracts Energy**: Describes the function of the turbine, which is to take energy from a fluid flow.
- **Fluid Flow**: The source of energy for the turbine.
- **Converts it into Useful Work**: Indicates the purpose of the turbine, which is to convert the extracted energy into useful work.
Can LLM do text augmentation?

Original sentence: The use of a cation cross-linked polysaccharide coating for reducing the glycemic response of a carbohydrate-containing food.

Traditional method
(NLP + Semantics Related Extraction)

- The process of applying a cation cross-connected pectin coating for diminishing the glycemic response of a carbohydrate-containing nourishment product.
- The utilize of an ion cross-connected pectin covering for diminishing the glycemic response of a carbohydrate-containing foodstuff.
- The application process of an ion cross-connected alginate film for lessening the glycemic response of a carbohydrate-containing nourishment product.

Not all part of a sentence can be substituted in text augmentation, since it changes the theme and rheme need to be kept intact to not change to underlying semantic the sentence

Using chatGPT 3.5 for paraphrasing

- Implementing a cation cross-linked polysaccharide coating to mitigate the glycemic response of carbohydrate-rich food.
- Utilizing a coating of cation cross-linked polysaccharide to minimize the glycemic impact of carbohydrate-containing food. (response and impact is not equivalent)
- Employing a cation cross-linked polysaccharide coating to decrease the glycemic response associated with a carbohydrate-based food.
Can LLM model do claim analyses comparison?

US2021350229A1
TRAINING TEXT SUMMARIZATION NEURAL NETWORKS
WITH AN EXTRACTED SEGMENTS PREDICTION

OBJECTIVE

What is claimed is: 1. A method comprising:

pre-training a text summarization neural network including learning values of a plurality of network parameters through self-supervised learning using unlabeled data comprising unlabeled first texts, the pre-training comprising:

obtaining an unlabeled first text comprising a plurality of segments;

selecting one or more of the plurality of segments;

processing, using the text summarization neural network and in accordance with current values of the plurality of network parameters, a masked first text that excludes the one or more selected segments to generate a prediction of the one or more selected segments; and

determining, based on a difference between the prediction and the one or more selected segments, an update to the current values of the plurality of network parameters; and

adapting the pre-trained text summarization neural network for a specific text summarization task including adjusting learned values of the plurality of network parameters using labeled data comprising second texts and respective summaries of the second texts.

US5918240A Automatic method of extracting summarization using feature probabilities

What is claimed is:

A processor implemented method of automatically extracting a subset of sentences from sentences of a natural language document presented in machine readable form to the processor, the document including a second multiplicity of sentences, the processor being coupled to a memory storing machine readable instructions for extracting sentences, the method comprising the steps of:

a) designating a sentence of the document as a selected sentence;

b) determining values for the selected sentence of each feature of a feature set, the feature set including a location feature and an upper case feature, the location feature having a first location value, a second location value, and a third location value, the first location value indicating that the selected sentence is included within a beginning portion of the document, the second location value indicating that the selected sentence is included within a middle portion of the document, and the third location value indicating that the selected sentence is included within an ending portion of the document, the upper case feature having a first upper case value and a second upper case value, the first upper case value indicating that selected sentence does not include any of a multiplicity of selected upper case phrases, the second upper case value indicating the selected sentence includes a one of the selected upper case phrases;

c) for each feature increasing a score for the selected sentence based upon the value of the feature for the selected sentence and upon a probability associated with the value of the feature;

d) if all sentences of the document have not been designated as the selected sentence, repeating steps a) through c); and

e) selecting the subset of sentences to be extracted based upon sentence scores.
Text 1 and Text 2 are different in terms of their content and the subject matter they describe:

Subject Matter:
- Text 1 is describing a method related to pre-training a text summarization neural network and adapting it for a specific text summarization task.
- Text 2 is describing a method for automatically extracting a subset of sentences from a natural language document based on features and scores.

Specificity:
- Text 1 is more specific in terms of its context, focusing on the pre-training and adaptation of a text summarization neural network.
- Text 2 is more general and seems to be about a method for sentence extraction from a natural language document using features and scores.

Technical Details:
- Text 1 includes technical details related to neural networks, parameter learning, and self-supervised learning.
- Text 2 includes technical details about features, probabilities, and scoring for sentence extraction.

Application:
- Text 1 is more related to natural language processing and text summarization tasks.
- Text 2 is more related to information retrieval or document processing tasks.

General analyses
- In summary, Text 1 is focused on neural network training and adaptation for text summarization, while Text 2 deals with the extraction of sentences from a document based on features and scores. They differ in their subject matter and technical details.
How do we solve the puzzle to automatic generate patent summaries?
What is the information requirement of good patent summarization
No 37435 Benz Patent – Motorwagen (1886)

Search for:
• Problem and Solution (A)
• Scope of the invention (A,B,C)
• Specific technical details (B,C)

The transport vehicle

Steering mechanism

Engine function
Discourse Information

• Lexical cohesion (TEXTTILIG)

• Connecting claim sentence to relevant text segment within the patent itself
  • Mase H., Matsubayashi T, Ogawa Y, Iwayama M, and Oshio T. Proposal of two-stage patent retrieval method considering the claim structure. 4(2):190–206, June 2005

• Sentence labelling
  • https://github.com/nlpTRIZ
  • Contradiction
  • https://github.com/Renuk9390/Patent_Sentiment_Analysis
    • Solution to problem
    • Advantageous effect invention
    • Technical problem
But most important are the **Technical Terms**

*The majority of entities in technical English dictionaries consist of terms with more than one word.*


*The identification of technology terms within a collection of patents is a challenging information extraction task due to the nature of technology terms themselves, which may be ambiguous or generic and have multiple nuances of interpretation.*

What is Technical Term in patent?
Depends on who you are asking and what context!

<table>
<thead>
<tr>
<th>Candidate Term</th>
<th>Word2Vec</th>
<th>C-Value</th>
<th>Pointwise Mutual information</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote communication</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Communication link</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Resin particle</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>washed washing</td>
<td>No/Yes (0.642)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bar code</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Wet strength</td>
<td>Not</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Our assumption

*Phrases having a homogenous distribution of IPC codes will reflect the termhoodness compared to phrases with heterogeneous distribution*
Using BERT – Technical Term Prediction

• Task: detect domain term spans in sentences

A network synchronization method allows

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>O</td>
<td></td>
</tr>
</tbody>
</table>

• Using pre-trained SciBERT Model
  • With additional training data for Technical Term detection
  • Randomly picked 22 patents with different International Patent Classification codes
  • Sentences 10,337, Technical Term dictionary of size 5,099

Fink T., Andersson L., Hanbury A. (2021) Detecting Multi Word Terms in patents the same way as entities / World Patent Information, 67 102078; 1 - 6
# Technical term identification
- Evaluation using WIPO PEARL Terminology DB

## Recall and Precision Assessment

<table>
<thead>
<tr>
<th>Term Category</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIPO Concept Recall</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>All Domain Term Recall</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>All Domain Term Precision</td>
<td>85%</td>
<td></td>
</tr>
</tbody>
</table>

### Domain-wise Recall and Precision

- **ADM & BANKING**: Recall 93%, Precision 88%
- **AERO/AERONAUTICS**: Recall 67%, Precision 70%
- **AGRI/AGRICULTURE**: Recall 73%, Precision 68%
- **BLDG/CIVIL ENGINEERING**: Recall 94%, Precision 86%
- **CHEM/CHMICAL**: Recall 65%, Precision 78%
- **DATA/COMPUTER SCIENCE**: Recall 65%, Precision 78%
- **ELEC/ELECTRICAL & R/C**: Recall 77%, Precision 66%
- **ENVR/ENVIRONMENTAL**: Recall 71%, Precision 80%
- **FOOD/FOOD TECHNOLOGY**: Recall 97%, Precision 88%
- **HOME/HOME TECHNOLOGY**: Recall 56%, Precision 94%
- **HORD/LEGAL MATTERS**: Recall 94%, Precision 92%
- **LEGAL**: Recall 73%, Precision 80%
- **MANU/MANUFACTURING**: Recall 72%, Precision 94%
- **MEAS/STANDARDIZATION**: Recall 95%, Precision 88%
- **MECH/MEDICAL SCIENCE & BIOLOGY**: Recall 73%, Precision 80%
- **METL/METALLURGY**: Recall 72%, Precision 82%
- **MINF/MINING**: Recall 97%, Precision 92%
- **PACK/PACKAGING & PAPER**: Recall 86%, Precision 90%
- **RALT/RAIL TRAVEL**: Recall 98%, Precision 92%
- **SCIE/SCIENCE**: Recall 98%, Precision 90%
- **SPRT/SPORTS, RECREATION & LEISURE**: Recall 98%, Precision 90%
- **TEXT/TEXTILE**: Recall 98%, Precision 90%
Why not combine the two?

Using semantically enriched data can retrieve up to 60% more related concepts than traditional pre-trained contextual models (a.k.a. Neural Network-based methods, BERT-family).

Artificial Researcher’s text mining technology

<table>
<thead>
<tr>
<th>Technical terms</th>
<th>Related concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>brake pedal:</td>
<td>vehicle operating pedal, conventional hydraulic brake system pedal devices position brake actuating member brake actuating member hydraulically-assisted rack pinion steering gear brake operating member conventional braking system pair pedals</td>
</tr>
<tr>
<td>accelerator pedal</td>
<td>case pedal device pedal device</td>
</tr>
</tbody>
</table>

Extraction using pre-trained contextual models only

<table>
<thead>
<tr>
<th>brake pedal</th>
<th>PatBERT</th>
<th>SciBERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>conventional hydraulic brake system</td>
<td>0.69</td>
<td>0.89</td>
</tr>
<tr>
<td>hydraulically-assisted rack pinion steering gear</td>
<td>0.49</td>
<td>0.80</td>
</tr>
<tr>
<td>conventional braking system</td>
<td>0.66</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Next step integrate the Domain Name Entity labels
Medical industry: population (patient group) and therapy

• The promise of extracorporeal liver replacement therapy and non-invasive ventilation were other areas of interest.
  • Therapy: extracorporeal liver replacement therapy

• Based on the first 35 patients admitted to the hospital with COVID-19, we evaluated the various symptoms with which patients presented
  • Population: patients admitted to the hospital with COVID-19
The promise of extracorporeal liver replacement therapy and non-invasive ventilation were other areas of interest.
Identification of fragrance concepts in text

A task-oriented BERT model can extract following entities form text

- Perfumes **smell** creamy due to **large doses** of **vanillic**, **musk**y and **milky** notes. **Coty Vanilla Fields** is a **familiar creamy vanilla fragrance**. An **opposite** of that is a **dry and sharp sensation**, similar to the one produced by **amber** in the base of **Paco Rabanne Black XS**.

**Smell** – different word and phrases related to fragrance concept related to fragrance referring to type

**Fragrance description** – A phrase or word describing type of fragrance

**Quantity** – Dosage and measures

**Substance entity** – Word or phrases defined as substance used for producing fragrance

**Adjective (descriptive & degree)** – Describing a type of relations, here adjective

**Proper name** – a name of perfume or brand
Why not join all three?
Text classification & semantic relation extraction & domain NE labelling

Input

Output

Automatic classification according to taxonomies such as IPC/CPC, MeSH, PLoS, WIPO Scientific Subject fields
So, to the question can we use LLM for patent summarization?
The current LLM

• For text argumentation

• As part of the verification of sentence categorization and NE labelling

• To make the extracted text more fluently
The initial patent summarization experiment was conducted at **PatentSemTech 2023**

- We assumed that participants of PatentSemTech were all knowledgeable in the field of text mining, so we put automatic patent summarization with an LLM (ChatGPT 3.5) to the test.

- We selected five patents within the field of text mining
  - For each of them, we have created four summarizations, one of which is manually curated by domain experts.

- We asked the participants to indicate their level of knowledge:
  - I have worked on related topics and have a good general understanding of the area of technology
  - I read few patents on this topic but not my main area of expertise
  - I read very few patents on this topic, and it is outside of my area of expertise

- We asked the participants to rank the summarizations from poor to excellent quality.
A method for summarizing text (20), comprising evaluating (24) selected words of the text according to predetermined criteria to provide word score values for each of the selected words. The method then provides for calculating (25) for each of the selected words a word weighted score that is dependent on the word score values and a number of occurrences of each of the selected words. Thereafter a step (26) of scoring sentences of the text to determine a sentence weighted score for the sentences is conducted. The sentence weighted score depends on sentence type and a combined word weighted score for words in the sentence. The method then provides for selecting (27) sentences to provide a summary of the text, the selecting being dependent on the sentence weighted score of the sentences.

**Link:** https://patents.google.com/patent/US20060206806
Preliminary results for two patents

Human versus LLM (chatGPT 3.5)

<table>
<thead>
<tr>
<th>US20060206806A1</th>
<th>Human</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Fair</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Excellent</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>US10831345B2</th>
<th>Human</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Fair</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Good</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Excellent</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

The estimated time for assessment by participants

<table>
<thead>
<tr>
<th>Time</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorter</td>
<td>11</td>
</tr>
<tr>
<td>15 to 25 minutes</td>
<td>4</td>
</tr>
<tr>
<td>25 to 35 minutes</td>
<td>0</td>
</tr>
<tr>
<td>longer</td>
<td>0</td>
</tr>
</tbody>
</table>
Questions!

Artificial Researcher IT GmbH is a Spin-off from TU Wien since 2019
More information visit https://artificialresearcher.com
Appendix
Artificial Researcher - Services & Products

Test access: API key b10225791d2a478c9ff3d8cd106ad65a

- **Artificial Researcher Passage Retrieval Service**
  - Access: rest API, Desktop App Script (Python)
    - [https://swagger.artificialresearcher.com/](https://swagger.artificialresearcher.com/) (developer)
    - [https://passageretrieval.artificialresearcher.com/](https://passageretrieval.artificialresearcher.com/) (demo)
  - SaaS or on-premises software (Docker)

- **Artificial Researcher Ontology Service**
  - Access: rest API, Desktop App Script (Python)
    - [https://swagger.artificialresearcher.com/?urls.primaryName=onto-api](https://swagger.artificialresearcher.com/?urls.primaryName=onto-api) (developer)
  - Data format OWL or JSON
  - SaaS or on-premises software (Docker)
    - Ontology generator – only SaaS

- **Artificial Researcher NLP-toolkit Services**
  - Access: rest API
    - [https://swagger.artificialresearcher.com/?urls.primaryName=unified-nlp-server](https://swagger.artificialresearcher.com/?urls.primaryName=unified-nlp-server) (developer)
  - SaaS or on-premises software (Docker)

- **Artificial Researcher Graph Search Service**
  - UX: [https://graph-demo.artificialresearcher.com/](https://graph-demo.artificialresearcher.com/) (demo)
The experiment

• For each patent, you will be presented with three summarizations created by algorithms, and one summary curated by a domain expert. The summarizations range between 250 to 700 words.

Instruction:

• Read the abstract and take a brief look at the patent using the provided link.
  • Please aim to spend no more than 15 minutes on reviewing the patent application itself.
• Grade the four different summarizations in terms of quality, ranging from poor to excellent.
  • You have the option to rate all of them as poor if none of them offer a representative summary of the invention.
Text summarization, US20060206806A1

https://forms.office.com/e/aEyR2tampW

Abstract
A method for summarizing text (20), comprising evaluating (24) selected words of the text according to predetermined criteria to provide word score values for each of the selected words. The method then provides for calculating (25) for each of the selected words a word weighted score that is dependent on the word score values and a number of occurrences of each of the selected words. Thereafter a step (26) of scoring sentences of the text to determine a sentence weighted score for the sentences is conducted. The sentence weighted score depends on sentence type and a combined word weighted score for words in the sentence. The method then provides for selecting (27) sentences to provide a summary of the text, the selecting being dependent on the sentence weighted score of the sentences.

Link: https://patents.google.com/patent/US20060206806
Unsupervised ontology-based graph extraction from texts, US10169454B2

Abstract

A method for extracting a relations graph uses an ontology graph in which nodes represent entity classes or concepts and edges represent properties of the classes. A property is associated with a constraint which defines a range of values that can be taken without incurring a cost. Input text in which entity and concept mentions are identified is received. An optimal set of alignments between a subgraph of the ontology graph and the identified mentions is identified by optimizing a function of constraint costs incurred by the alignments and a distance measure computed over the set of alignments. The relations graph is generated, based on the optimal set of alignments. The relations graph represents a linked set of relations instantiating a subgraph of the ontology. The relations graph can include relations involving implicit mentions corresponding to subgraph nodes that are not aligned to any of the concept or entity mentions.

Link: https://patents.google.com/patent/US10169454B2
Establishing user specified interaction modes in a question answering dialogue, US10831345B2

https://forms.office.com/e/VTZ3LY02Lm

Abstract

An approach is provided for automatically generating user-specific interaction modes for processing question and answers at the information handling system by receiving a question from a user, extracting user context parameters identifying a usage scenario for the user, identifying first input and output presentation modes for the user based on the extracted user context parameters, monitoring user interaction with the system in relation to the question, and adjusting the first input and output presentation modes based on the extracted user context parameters and detected user interaction with the system. Link: https://patents.google.com/patent/US10831345B2
Title: User-centric soft keyboard predictive technologies

Abstract

An apparatus and method are disclosed for providing feedback and guidance to touch screen device users to improve text entry user experience and performance by generating input history data including character probabilities, word probabilities, and touch models. According to one embodiment, a method comprises receiving first input data, automatically learning user tendencies based on the first input data to generate input history data, receiving second input data, and generating auto-corrections or suggestion candidates for one or more words of the second input data based on the input history data. The user can then select one of the suggestion candidates to replace a selected word with the selected suggestion candidate.

Title: Regularized latent semantic indexing for topic modeling, US8533195B2

Abstract

Electronic documents are retrieved from a database and/or from a network of servers. The documents are topic modeled in accordance with a Regularized Latent Semantic Indexing approach. The Regularized Latent Semantic Indexing approach may allow an equation involving an approximation of a term-document matrix to be solved in parallel by multiple calculating units. The equation may include terms that are regularized via either l1 norm and/or via l2 norm. The Regularized Latent Semantic Indexing approach may be applied to a set, or a fixed number, of documents such that the set of documents is topic modeled. Alternatively, the Regularized Latent Semantic Indexing approach may be applied to a variable number of documents such that, over time, the variable of number of documents is topic modeled.

Link: https://patents.google.com/patent/US8533195B2
Artificial Researcher - Services & Products

Test access: API key b10225791d2a478c9ff3d8cd106ad65a

- **Artificial Researcher Passage Retrieval Service**
  - Access: rest API, Desktop App Script (Python)
    - [https://swagger.artificialresearcher.com/](https://swagger.artificialresearcher.com/) (developer)
    - [https://passageretrieval.artificialresearcher.com/](https://passageretrieval.artificialresearcher.com/) (demo)
  - SaaS or on-premises software (Docker)

- **Artificial Researcher Ontology Service**
  - Access: rest API, Desktop App Script (Python)
    - [https://swagger.artificialresearcher.com/?urls.primaryName=onto-api](https://swagger.artificialresearcher.com/?urls.primaryName=onto-api) (developer)
    - Data format OWL or JSON
  - SaaS or on-premises software (Docker)
    - Ontology generator – only SaaS

- **Artificial Researcher NLP-toolkit Services**
  - Access: rest API
    - [https://swagger.artificialresearcher.com/?urls.primaryName=unified-nlp-server](https://swagger.artificialresearcher.com/?urls.primaryName=unified-nlp-server) (developer)
  - SaaS or on-premises software (Docker)

- **Artificial Researcher Graph Search Service**
  - UX: [https://graph-demo.artificialresearcher.com/](https://graph-demo.artificialresearcher.com/) (demo)