

# Generating Acrostics via Paraphrasing and Heuristic Search

Bruno Soares Fillmann  
Fernando Bombardelli da Silva  
Jürgen Bauer  
William Bombardelli da Silva

Technische Universität Berlin  
Datenbanksysteme und Informationsmanagement  
DBPRO – Database Projects (WS 2014/2015)

12.01.2015

# Organization

- 1 Overview
- 2 New Operations
- 3 New Results
- 4 Conclusion
- 5 References

# Overview

- **The problem:** Given a text  $T$  and an acrostic  $X$ , find a paraphrased version of  $T$  that encodes  $X$  in the first letters of each line.
- **Main goal of the project:** Implement the paper for the German language.
- **Main idea of the algorithm:**
  - Modeled as a search problem in a tree.
  - The vertices are states (texts) and the edges are operations over states.
  - Artificial intelligence is applied for the search strategy ( $A^*$  Algorithm).

# Last State of Work

- Line Break and Wrong Hyphenation Working Properly
- Search Algorithm Working Properly
- Word Insertion Working Partially, but Deletion Not Working

# New Developed Operations

- Word Insertion Or Deletion (Enhancements)
- Synonym
- Hyphenation
- Spelling

# Synonym

- Free dictionary of synonyms is used, namely, Open Thesaurus ([www.openthesaurus.de](http://www.openthesaurus.de))
- We use **Redis** as database server (NoSQL)
- Redis stores data as key-value pairs
- Our base is structured in a way that, every word in the thesaurus is a key that points a set of synonyms
- The thesaurus text file is imported into the database by a Python script
- The database is then consulted by the application during its execution

# Hyphenation

- Line Length constraints  $l_{min} = 50$  and  $l_{max} = 70$ .
- For hyphenation a re-implementation of Knuth's hyphenation algorithm in TeX is used (TEXHyphenator-J by David Tolpin)
- A word can be hyphenated if: The part of the current line **from** the beginning **to** the hyphen has a length of at least  $l_{min} = 50$ .
- The text **after** the hyphen has to be aligned again to fulfill the line length constraints.
- A greedy word wrap algorithm is applied.
- Don't allow words of length  $> 20$  in the start text.

# Hyphenation

- **Example:**
- 38 hyphenations!

ORIGINAL TEXT:

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium .....

0.

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium .....

1.

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium .....

2.

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium .....



# Results - Synonym

Deutsches Leben der Gegenwart -- dem feindlichen Blick, der nur seine Oberfläche streift, möchte scheinen, daß die Gegenwart wenig vom deutschen Leben, mehr vom deutschen Sterben zu melden hätte. Aber der nachdenkliche Betrachter weiß, daß die größten geistigen Epochen Deutschlands über seinen politischen Niederlagen wuchsen, daß gerade die Zeiten nach dem Dreißigjährigen Krieg, nach dem Zusammenbruch von Jena zu den schöpferischen des deutschen Lebens gehören.

Deutsches Leben der Gegenwart -- dem feindlichen Blick, der gerade  
eben seine Oberfläche streift, möchte scheinen, daß die Gegenwart  
unbedeutend vom deutschen Leben, mehr vom deutschen Sterben zu melden hätte.  
trotzdem der nachdenkliche Betrachter weiß, daß die größten geistigen  
Epochen Deutschlands über seinen politischen Niederlagen wuchsen, daß  
nachgerade die Zeiten nach dem Dreißigjährigen Krieg, nach dem  
Zusammenbruch von Jena zu den schöpferischen des deutschen Lebens  
gehören.

// Synonym  
// Line break  
// Synonym  
// Synonym  
// Line break  
// Synonym

# Results – Deletion Of Words

In diesen Auflagen könnte denn auch der Hebel liegen, um Griechenland weiteres Geld zu verweigern. Denn ein Stopp der Zinszahlungen dürfte als Verstoß gegen das Rettungsprogramm gewertet werden. Denn für alle Geschäftsbanken in der Eurozone gilt: Geld von der EZB gibt es nur mäßig, wenn die Finanzinstitute eine bestimmte Menge an Wertpapieren, zum Beispiel Staatsanleihen, als Pfand bei der EZB hinterlegen. In diesem Fall könnte die EZB ihre Ansprüche an die Sicherheiten wieder erhöhen.

## Result Text:

In diesen Auflagen könnte denn auch der Hebel liegen, um Griechen-  
nland weiteres Geld zu verweigern. Denn ein Stopp der Zinszahlungen dürf-  
te als Verstoß gegen das Rettungsprogramm gewertet werden. Denn für alle  
in der Eurozone gilt: Geld von der EZB gibt es nur  
mäßig, wenn die Finanzinstitute eine bestimmte Menge an Wertpapieren,  
zum Beispiel Staatsanleihen, als Pfand bei der EZB hinterlegen. In  
diesem Fall könnte die EZB ihre Ansprüche an die Sicherheiten wieder  
erhöhen.

*//Wrong Hyphenation*

*//Wrong Hyphenation*

*//Word Deletion*

# Results — Insertion Of New Words

Im Englischen bedeutet blau sein so etwas wie traurig oder deprimiert sein, weltbekannt ist die Musikrichtung Blues. Das ist traurig aber. Im Deutschen hat blau sein allerdings eine ganz andere Bedeutung. Wenn jemand eine feucht-fröhliche Party feiert (also eine, auf der viel Alkohol getrunken wird), dann ist er hinterher wahrscheinlich blau. Blau sein bedeutet im Deutschen nämlich betrunken sein.

## Result Text:

Im Englischen bedeutet blau sein so etwas wie traurig oder deprimiert sein, weltbekannt ist die Musikrichtung Blues. Das ist echt traurig aber. Im Deutschen hat blau sein allerdings eine ganz andere Bedeutung. Wenn jemand eine feucht-fröhliche Party feiert (also eine, auf der viel Alkohol getrunken wird), dann ist er hinterher wahrscheinlich blau. Blau sein bedeutet im Deutschen nämlich betrunken sein.

*//Line Break*

*//Word Insertion*

*//+ Line Break*

# Examples

- 1 min 22 sec, 17517 nodes, Ops: Hyph,WrHyph,LB

## ORIGINAL TEXT:

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover. (WrongHyphenation)

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover. (WrongHyphenation)

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover. (Hyphenation)

# Examples

- 1 min 22 sec, 17517 nodes, Ops: Hyph,WrHyph,LB

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover. (Hyphenation)

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover. (LineBreak)

Friedrich wurde im Berliner Stadtschloss geboren und war der älteste überlebende Sohn von insgesamt 14 Kindern König Friedrich Wilhelms I. und dessen Gattin Sophie Dorothea, der ersten hochadligen Königin des Fürstentums zu Hannover.

# Conclusion

- Present operations
- Update in the state of the work (what can the algorithm reach, relation with the paper)
- Difficulties:
  - The time consumed by the requests to the web service
  - The choice of the quality of each operation is decisive in the result
  - ...

# References



Benno Stein, Matthias Hagen, and Christof Bräutigam. *Generating Acrostics via Paraphrasing and Heuristic Search*.

In Junichi Tsujii and Jan Hajic, editors, 25th International Conference on Computational Linguistics (COLING 14), pages 2018-2029, August 2014. Association for Computational Linguistics.



Spiegel Online. *Griechenland und die Euro-Zone: Der fast unmögliche Rausschmiss*.

<http://www.spiegel.de/wirtschaft/soziales/griechenland-euro-austritt-waere-moeglich-aber-kompliziert-a-1011361.html>. Am 07. Januar 2015



Hilko. *Blau ist keine traurige Farbe*.

<https://deutschlich.wordpress.com/2013/03/12/blau-ist-keine-traurige-farbe>. Am 07. Januar 2015



Redis. <http://redis.io/>. Am 07. Januar 2015



Synonym - *OpenThesaurus - Deutscher Thesaurus*.

<https://www.openthesaurus.de/>. Am 07. Januar 2015

# Questions?