

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)

01.12.2014

Organisation

- 1 Overview
- 2 Implemented Operations
- 3 First Results
- 4 Next Steps
- 5 References

Overview

- **Entrada:** Um grafo virtual e um grafo do substrato físico
- **Problema de Localização:** Encontrar um mapeamento *válido* dos vértices do grafo virtual para os vértices do grafo físico, assim como das arestas do virtual para caminhos no físico
- **Problema de Otimização:** Minimizar a banda total utilizada pelo mapeamento

Classes Diagram

Activity Diagram and Search Strategy

Line Break

- Constraints on line length $l_{min} = 50$ and $l_{max} = 70$.
- Two cases:
 - After a word when line length falls in the $[l_{min}, l_{max}]$ -window.
 - After a full stop. (i.e., paragraph break)
- Succeeding line break, lines have to be aligned.
- A greedy word wrap algorithm is applied.
- Avoid words of length > 20 in the start text.

Line Break

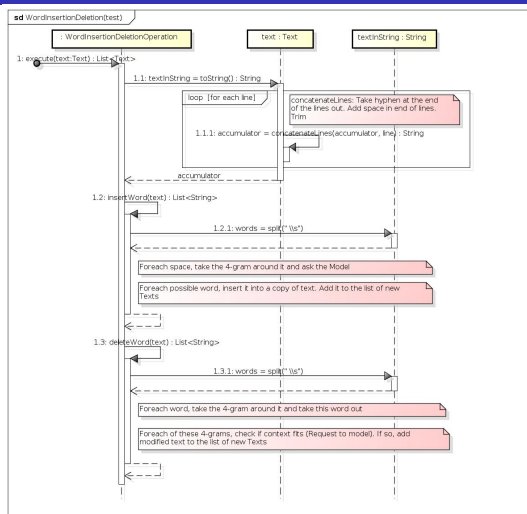
- **Example:**
- 18 linebreaks,
- No full stop after 28!
- Third often applied operator on a solution path
- Fastest operator

Johann Wolfgang von Goethe wurde am 28. August 1749 im heutigen Goethe-Haus am Frankfurter Großen Hirschgraben geboren. Der Vater Johann Caspar Goethe (1710–1782), der die ursprüngliche Schreibung des Familiennamens von Göthe in Goethe änderte, war promovierter Jurist, übte diesen Beruf jedoch nicht aus, sondern lebte von den Erträgen seines ererbten Vermögens, das später auch dem Sohn ein Leben ohne finanzielle Zwänge ermöglichen sollte. Er war vielseitig interessiert und gebildet, jedoch auch streng und pedantisch, was wiederholt zu Konflikten in der Familie führte.

Wrong Hyphenation

Ensure Constraints

Word Insertion



Word Insertion – Databases

- Current: Netspeak Requests (*netspeak.org*) [2]
- Possibility: Google NGram Database Manually Handled

First Results – Donald Knuth

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium am California Institute of Technology im September 1956. Aus zweierlei Gründen schlug er tatsächlich seinem zweiten Studienjahr jedoch den Weg zur Mathematik ein: Zum einen löste er ein Problem eines seiner Mathematikprofessoren, was ihm eine 1,0 als Note einbrachte, zum anderen fand er wenig Gefallen an den physikalischen Praktika. Er erhielt einen Bachelor- und einen Master-Abschluss 1960 an der Case Western Reserve University. 1963 erhielt er seinen Ph.D. vom California Institute of Technology bei Marshall Hall, wo er dann auch nach der Promotion Assistant Professor und 1966 Associate Professor und schließlich Professor wurde. 1968 wurde er Professor für Informatik an der Stanford University.

Result Text:

Knuth ist der Sohn eines Lehrers. Er besuchte die Milwaukee Lutheran High School und begann sein Physikstudium am California Institute of Technology im September 1956. Aus zweierlei Gründen schlug er tatsächlich seinem zweiten Studienjahr jedoch den Weg zur Mathematik ein: Zum einen löste er ein Problem eines seiner Mathematikprofessoren, was ihm eine 1,0 als Note einbrachte, zum anderen fand er wenig Gefallen an den physikalischen Praktika. Er erhielt einen Bachelor- und einen Master-Abschluss 1960 an der Case Western Reserve University. 1963 erhielt er seinen Ph.D. vom California Institute of Technology bei Marshall Hall, wo er dann auch nach der Promotion Assistant Professor und 1966 Associate Professor und schließlich Professor wurde. 1968 wurde er Professor für Informatik an der Stanford University.

//Wrong Hyphenation
//Wrong Hyphenation
//Line Break
//Wrong Hyphenation

First Results — Goethe

Johann Wolfgang von Goethe wurde am 28. August 1749 im heutigen Goethe-Haus am Frankfurter Großen Hirschgraben geboren. Der Vater Johann Caspar Goethe (1710{1782), der die ursprüngliche Schreibung des Familiennamens von Göthe in Goethe änderte, war promovierter Jurist, übte diesen Beruf jedoch nicht aus, sondern lebte von den Erträgen seines ererbten Vermögens, das später auch dem Sohn ein Leben ohne finanzielle Zwänge ermöglichen sollte. Er war vielseitig interessiert und gebildet, jedoch auch streng und pedantisch, was wiederholt zu Konflikten in der Familie führte.

Result Text:

Johann Wolfgang von Goethe wurde am 28. August 1749 im he- *//Wrong Hyphenation*
 utigen Goethe-Haus am Frankfurter Großen Hirschgraben gebore- *//LineBreak + Wrong Hyphenation*
 n. Der Vater Johann Caspar Goethe (1710{1782), der die ursprüngl- *//Wrong Hyphenation*
 gliche Schreibung des Familiennamens von Göthe in Goethe änderte, war
 promovierter Jurist, übte diesen Beruf jedoch nicht aus, sondern lebte
 von den Erträgen seines ererbten Vermögens, das später auch dem Sohn
 ein Leben ohne finanzielle Zwänge ermöglichen sollte. Er war
 vielseitig interessiert und gebildet, jedoch auch streng und
 pedantisch, was wiederholt zu Konflikten in der Familie führte.

First Results — Berlin

Berlin ist die Hauptstadt und zugleich ein Land der Bundesrepublik Deutschland. Der Stadtstaat Berlin bildet eine Einheitsgemeinde und ist mit über 3,4 Millionen Einwohnern die bevölkerungsreichste und mit knapp 892 Quadratkilometern die flächengrößte Kommune Deutschlands sowie nach Einwohnern die zweitgrößte der Europäischen Union. Zudem ist Berlin mit rund 3800 Einwohnern je Quadratkilometer die am drittdichtesten bevölkerte Gemeinde Deutschlands. Berlin ist eine Enklave im Land Brandenburg und bildet das Zentrum der Metropolregion Berlin/Brandenburg (6 Millionen Einwohner) sowie der Agglomeration Berlin (4,4 Millionen Einwohner). Der Stadtstaat unterteilt sich in zwölf Bezirke.

Result Text:

Berlin ist die Hauptstadt und zugleich ein Land der Bundesrepublik Deutschland. Der Stadtstaat Berlin bildet eine Einheitsgemeinde und ist mit über 3,4 Millionen Einwohnern die bevölkerungsreichste und mit knapp 892 Quadratkilometern die flächengrößte Kommune Deutschlands sowie nach Einwohnern die zweitgrößte der Europäischen Union. Zudem ist Berlin mit rund 3800 Einwohnern je Quadratkilometer die am drittdichtesten bevölkerte Gemeinde Deutschlands. Berlin ist eine Enklave im Land Brandenburg und bildet das Zentrum der Metropolregion Berlin/Brandenburg (6 Millionen Einwohner) sowie der Agglomeration Berlin (4,4 Millionen Einwohner). Der Stadtstaat unterteilt sich in zwölf Bezirke.

//Wrong Hyphenation

//Line Break + Line Break

//Wrong Hyphenation

Next Steps

- Mapeamento de redes virtuais é complexo até mesmo para localização de solução inicial
- Importância do algoritmo de vizinhança para uma boa convergência do método
- *Solvers* genéricos não são viáveis na prática
- Decisões na fase de projeto do algoritmo são cruciais para o resultado com a meta-heurística



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.



Martin Potthast, Martin Trenkmann, and Benno Stein.
Netspeak: Assisting Writers in Choosing Words.
In Cathal Gurrin et al, editors, Advances in Information
Retrieval. 32nd European Conference on Information Retrieval
(ECIR 10) volume 5993 of Lecture Notes in Computer Science,
pages 672, Berlin Heidelberg New York, March 2010. Springer.

Questions?