



Introduction



Knuth-Plass



Extension



Experiments



Conclusion

Similarity Problems in Paragraph Justification: an Extension to the Knuth-Plass Algorithm

~ DocEng 2024 ~

Didier Verna

EPITA / LRE

didier@lrde.epita.fr



[lrde/~didier](#)



[@didierverna](#)



[didier.verna](#)



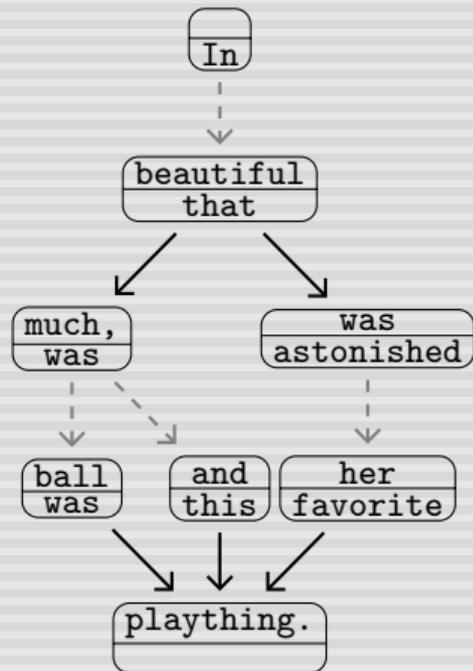
[in/didierverna](#)

Similarities

In olden times when wishing still helped one, there lived a king whose daughters were all beautiful; and the youngest was so beautiful that the sun itself, which has seen so much, was astonished whenever it shone in her face. Close by the king's castle lay a great dark forest, and under an old lime-tree in the forest was a well, and when the day was very warm, the king's child went out into the forest and sat down by the side of the cool fountain; and when she was bored she took a golden ball, and threw it up on high and caught it; and this ball was her favorite plaything.

- ▶ T_EX's rendition
 - ▶ Line width: 201pt
 - ▶ Font: Latin Modern Roman 10pt
- ▶ Problems (also at beginning of line)
 - ▶ Micro-interruption
 - ▶ Accidental line skipping
- ▶ Ancient: scribal errors (monks)
- ▶ Terminology
 - ▶ "Saut du même au même" (french) 🤔
 - ▶ Homeoarchy / homeoteleuton (rhetoric)
 - ▶ Character / word ladders ? 🤔

The Knuth-Plass in a Nutshell



- ▶ “Single Pair Shortest Path” problem
Finding the “best” route from beginning to end of paragraph
- ▶ Dynamic optimization technique
Never construct the full graph entirely
- ▶ Cost function
 - ▶ Local demerits: line by line
Badness, hyphenation penalty, etc.
 - ▶ Contextual demerits: line against line
Hyphenation ladders, stretching / shrinking disparities, etc.
- ▶ Decision: $\min \sum (\text{local} + \text{contextual demerits})$

Knuth-Plass Extension

- ▶ **Idea:** a new kind of contextual demerits (“similar demerits”)
 - ▶ Compare the beginning & end of consecutive lines
 - ▶ Add similar demerits to the total if identical
 - ▶ Note: special case for the final 2 lines
- ▶ **Implementation:**
 - ▶ remember bol / eol characters in nodes
 - ▶ only up to the first “glue” (elastic space) or “discretionary” (hyphenation point / ligature)
 - ▶ discard kerns
- ▶ **Rationale:**
 - ▶ Discretionaries: avoid deconstruction overhead
 - ▶ Glue: avoid vertical alignment considerations
 - ▶ Kerns: small adjustments, identical in similarities

Knuth-Plass Extension

▶ **Idea:** a new kind of contextual demerits (“similar demerits”)

- ▶ Compare the beginning & end of consecutive lines
- ▶ Add similar demerits to the total if identical
- ▶ Note: special case for the final 2 lines

▶ **Implementation:**

- ▶ rem **Comparing two short sequences of characters**
- ▶ only
- ▶ disc

He who can do less can do more...

int / ligature)

▶ **Rationale:**

- ▶ Discretionaries: avoid deconstruction overhead
- ▶ Glue: avoid vertical alignment considerations
- ▶ Kerns: small adjustments, identical in similarities

In olden times when wishing still helped one, there lived a king whose daughters were all beautiful; and the youngest was so beautiful that the sun itself, which has seen so much, was astonished whenever it shone in her face. Close by the king's castle lay a great dark forest, and under an old lime-tree in the forest was a well, and when the day was very warm, the king's child went out into the forest and sat down by the side of the cool fountain; and when she was bored she took a golden ball, and threw it up on high and caught it; and this ball was her favorite plaything.

Similar demerits = 0

In olden times when wishing still helped one, there lived a king whose daughters were all beautiful; and the youngest was so beautiful that the sun itself, which has seen so much, was astonished whenever it shone in her face. Close by the king's castle lay a great dark forest, and under an old lime-tree in the forest was a well, and when the day was very warm, the king's child went out into the forest and sat down by the side of the cool fountain; and when she was bored she took a golden ball, and threw it up on high and caught it; and this ball was her favorite plaything.

Similar demerits = 2800

In olden times when wishing still helped one, there lived a king whose daughters were all beautiful; and the youngest was so beautiful that the sun itself, which has seen so much, was astonished whenever it shone in her face. Close by the king's castle lay a great dark forest, and under an old lime-tree in the forest was a well, and when the day was very warm, the king's child went out into the forest and sat down by the side of the cool fountain; and when she was bored she took a golden ball, and threw it up on high and caught it; and this ball was her favorite plaything.

Similar demerits = 5230

Is it worth it?

- ▶ **Pertinence:** is the problem frequent?
- ▶ **Efficacy:** does this Knuth-Plass extension solve the problem?
- ▶ **Two typesetting experiments:**
 - ▶ One paragraph at many different widths
 - ▶ Many paragraphs at a single width

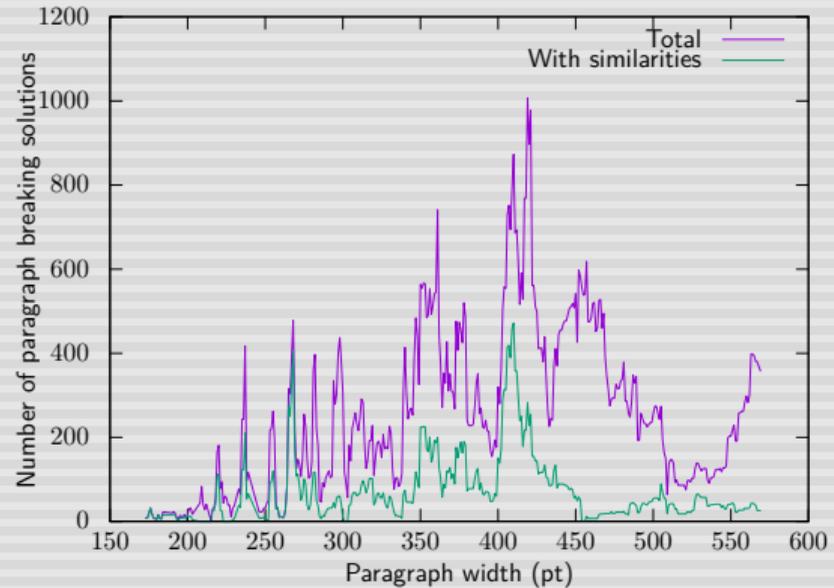
Pertinence

Experimental conditions

- ▶ Grimm Brothers “Frog King” §1
- ▶ 142pt ($\approx 5\text{cm}$) \rightarrow 569pt ($\approx 20\text{cm}$)
- ▶ 427 runs

Expected results

- ▶ Chaotic shapes
- ▶ Close numbers for narrow paragraphs



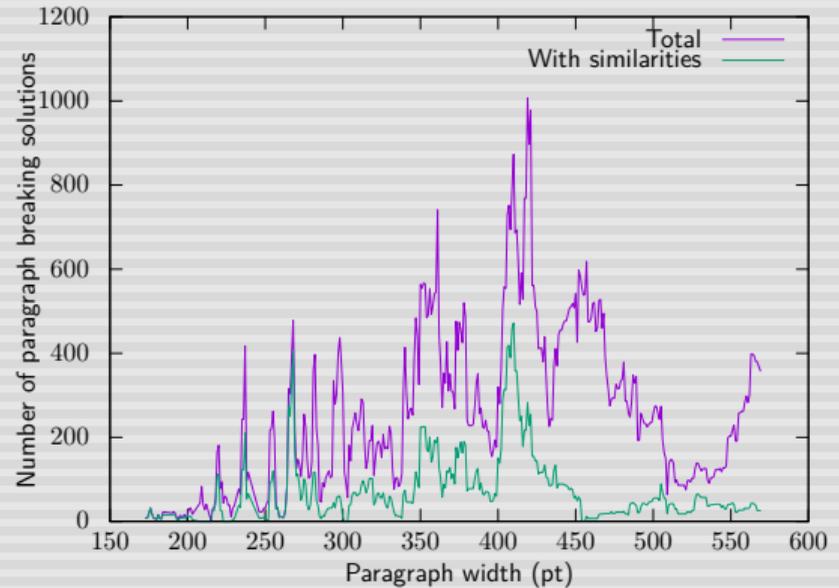
Pertinence

Experimental conditions

- ▶ Grimm Brothers “Frog King” §1
- ▶ 142pt ($\approx 5\text{cm}$) \rightarrow 569pt ($\approx 20\text{cm}$)
- ▶ 427 runs

Findings

- ▶ Similarities are frequent
- ▶ So are similarity-free layouts



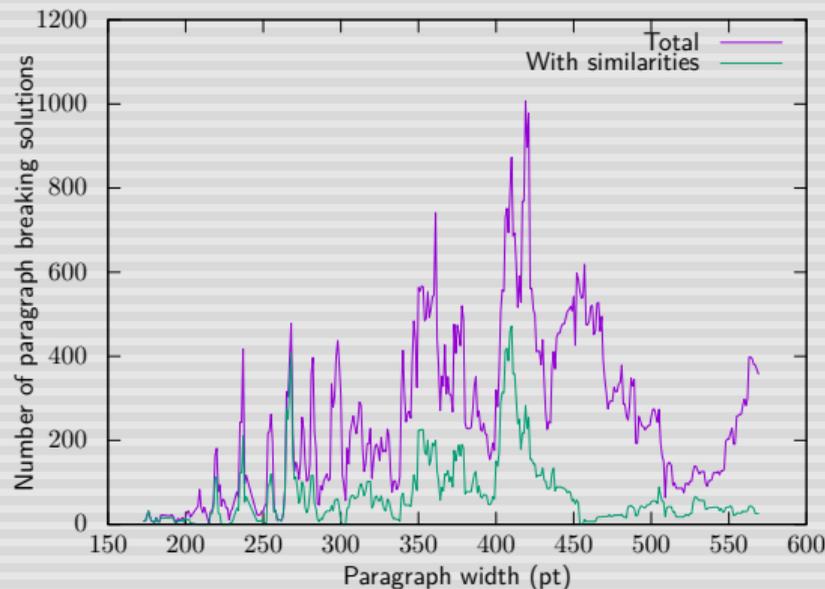
Pertinence

Experimental conditions

- ▶ Grimm Brothers “Frog King” §1
- ▶ 142pt ($\approx 5\text{cm}$) \rightarrow 569pt ($\approx 20\text{cm}$)
- ▶ 427 runs

What about $\text{T}_{\text{E}}\text{X}$?

- ▶ 4%: similarities unavoidable
- ▶ 21%: with similarities
- ▶ 26%: experiment 2



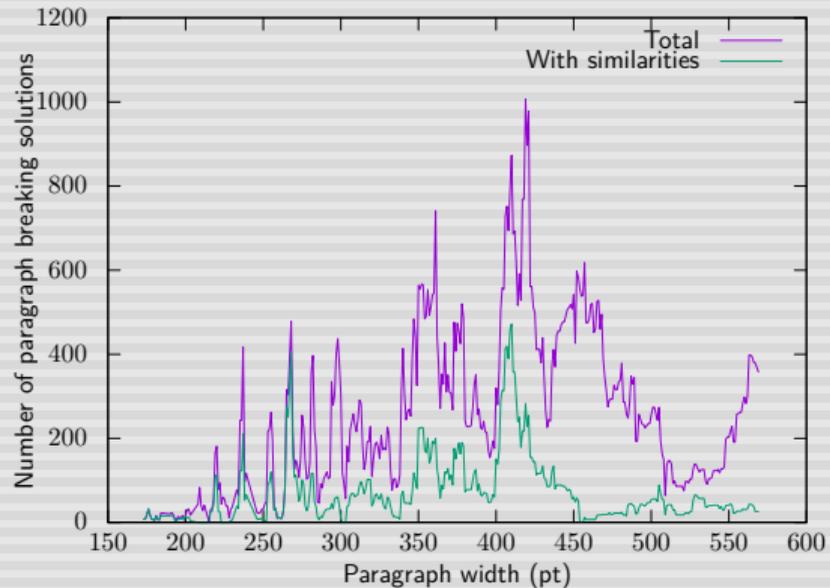
Pertinence

Experimental conditions

- ▶ Grimm Brothers “Frog King” §1
- ▶ 142pt ($\approx 5\text{cm}$) \rightarrow 569pt ($\approx 20\text{cm}$)
- ▶ 427 runs

Conclusion

- ▶ 2 paragraphs out of 10 is a lot
- ▶ The problem is worth addressing



Efficacy

Experimental conditions

- ▶ Herman Melville's Moby Dick
- ▶ 1524 paragraphs @ 284pt ($\approx 10\text{cm}$)
- ▶ + experiment 1 \rightarrow 1951 cases
- ▶ x3 \rightarrow 5853 runs

Conclusion

- ▶ Automatic treatment possible

Max. similar demerits	+ Disregard adjacency pb.
Corrected: 48% / 50%	Corrected: 53% / 66%
Improved: 50% / 63%	Improved: 57% / 73%

Please note

- ▶ Those are extreme conditions!
- ▶ Similarity-free \neq aesthetically better
- ▶ But we have some leeway...

Conclusion

- ▶ Similarity handling for the Knuth-Class
 - ▶ Simple & lightweight
 - ▶ Backward-compatible with regular $\text{T}_{\text{E}}\text{X}$ (similar demerits = 0)
- ▶ Implemented in ETAP, could be useful in production engines
- ▶ Experimentation \rightarrow automatic similarity handling is both worthy and achievable

Perspectives

- ▶ Study trade-off with other aesthetic criteria
- ▶ Similar demerits as $f(\textit{similarity})$ rather than just a scalar value
- ▶ LuaMeta $\text{T}_{\text{E}}\text{X}$ \rightarrow left/right “twin” demerits