DocuCompass: Effective Exploration of Document Landscapes
Heimerl et al. (2017)
Paper Analysis

Eric M. Cabral

Universidade de São Paulo (USP)
Instituto de Ciências Matemáticas e de Computação (ICMC)
Laboratório de Visualização, Imagens e Computação Gráfica (VICG)

cabral.eric@usp.br

June 19, 2018

Abstract

- ... free exploration and navigation on a level of abstraction between a labeled document spatialization and reading single documents is largely unsupported.
- DocuCompass, a focus+context approach based on the lens metaphor.
- It comprises multiple methods to characterize local groups of documents, and to efficiently guide exploration based on users' requirements.
- DocuCompass thus allows for effective interactive exploration of document landscapes without disrupting the mental map of users by changing the layout itself.

Keywords: interaction techniques, document visualization, text mining, visual analytics, focus+context

Table of Contents

Problem I

- Written natural languages are the most universal visual system of symbols that exists
  - Capable of encoding very complex information
- Using visualization to aggregate and analyze text is only a viable option, if the goal is to either abstract or summarize texts, or to extract and explicate very specific details
- Terms as labels
- Users' world knowledge and language comprehension skills are used as an integral part of analysis

Problem II

- Many visual methods for analyzing text are designed for information needs, analysis goals, or extraction tasks that are known in advance
  - User's previous knowledge
- Free exploration with no or very little prior knowledge about the corpus is hampered due to missing suitable interaction techniques

Solution I

- DocuCompass
  - Works with 2D document spatializations
  - Supports interactive abstraction and explication tasks on subsets of text documents
  - Coarse and fine-grained exploration down to the level of single documents is supported through interaction
  - Lens-based
  - focus+context technique

Solution II

- *"The technique we present constitutes an advancement for text exploration tasks by facilitating explorative analyses of large text collections. In addition, it offers navigation support to help users form and solidify an information need during initial, explorative analysis"*
- *"We offer an initial user study that indicates the effectiveness of our technique"*

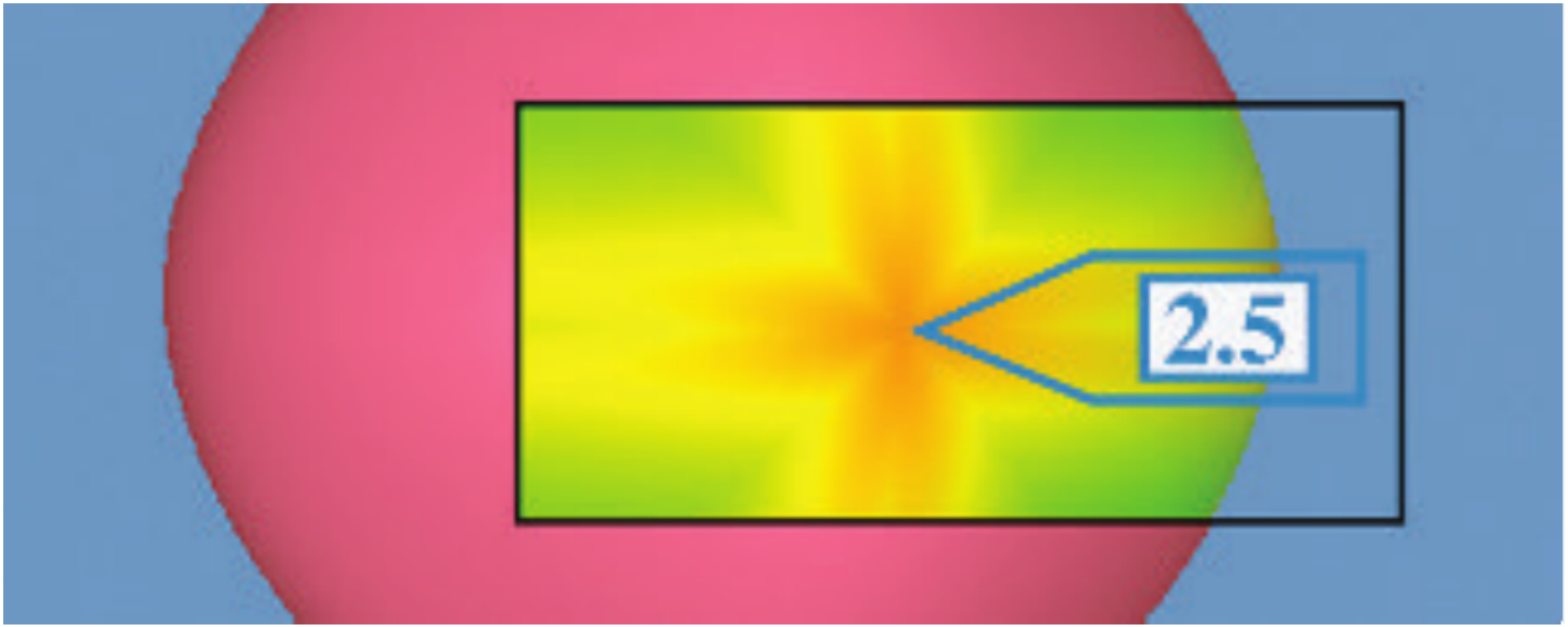
Focus+Context I

T. Munzner, *Visualization Analysis and Design*. A K Peters/CRC Press, 2014

- **Embed** detailed information about a selected set (**focus**) without losing the overview (**context**) of the visualization
  - Mitigate potential disorientation
  - Interaction
  - Filter and aggregation
  - Reduce internal cognitive load
- Elide
- **Superimpose**
- Distort

Focus+Context II

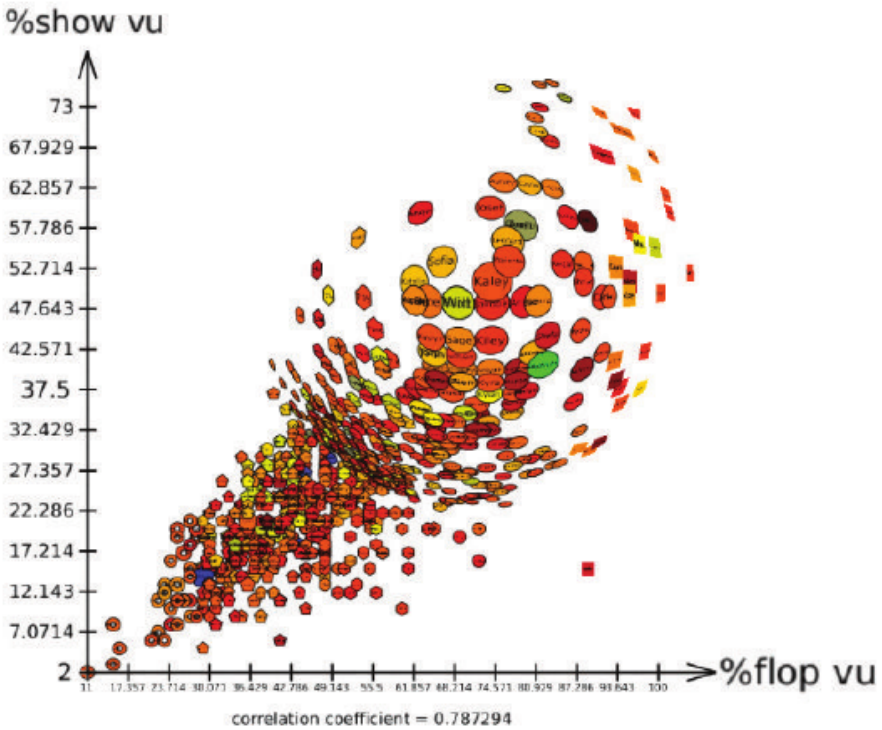
Figure: Magic lenses

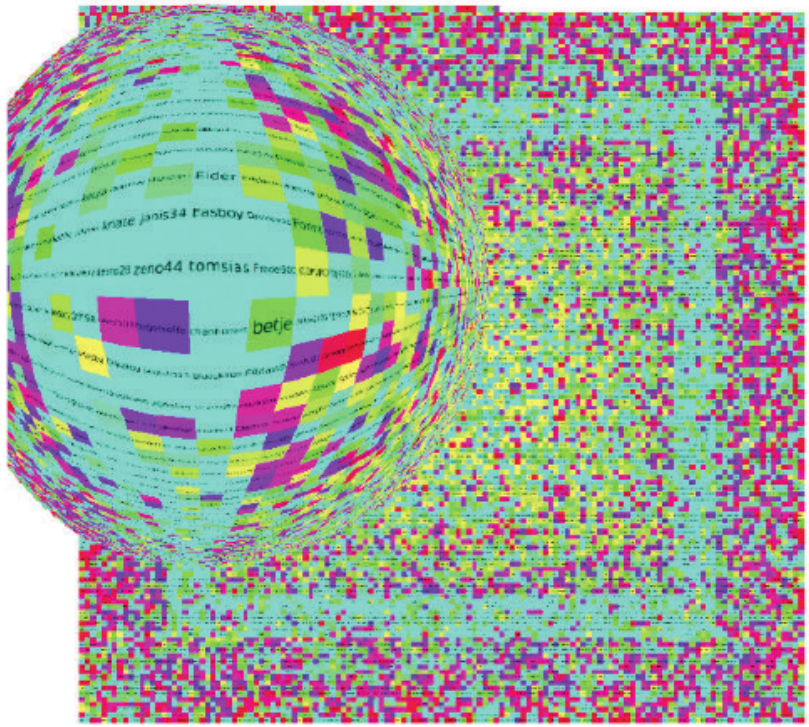


Source: Munzner (2014)

Focus+Context III

Figure: Distortion



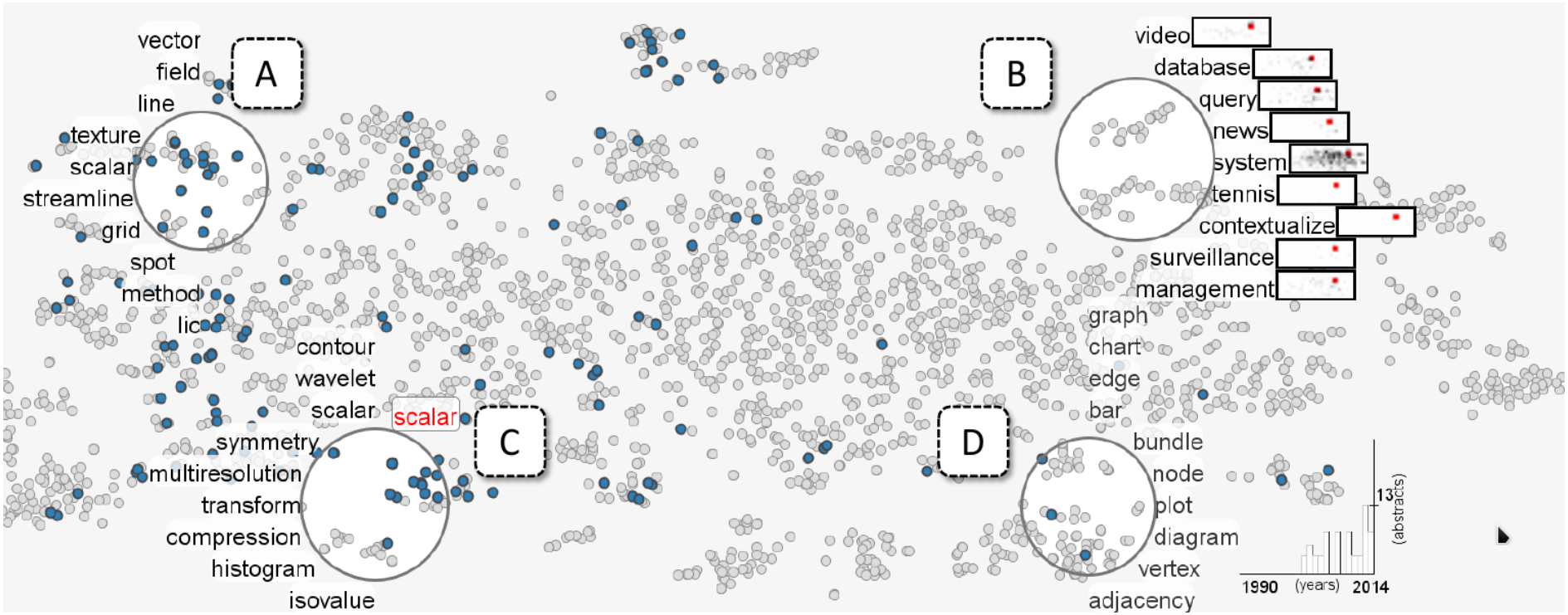


Source: Munzner (2014)

Hypothesis

- A magic lens-based technique has the benefit of keeping the context of the lens visually unchanged, or at least static with respect to the geometrical position of visual elements
- Lens-based techniques are therefore a natural fit for exploring text collections
- DocuCompass is a powerful interaction approach to complement traditional techniques which offer either overview or text details, and lack some intermediate interaction method

Figure: Overview



Source: Heimerl et al. (2017)

Table of Contents

Related Work

- Text Spatialization
- Focus+Context Interaction and Lenses

Table of Contents

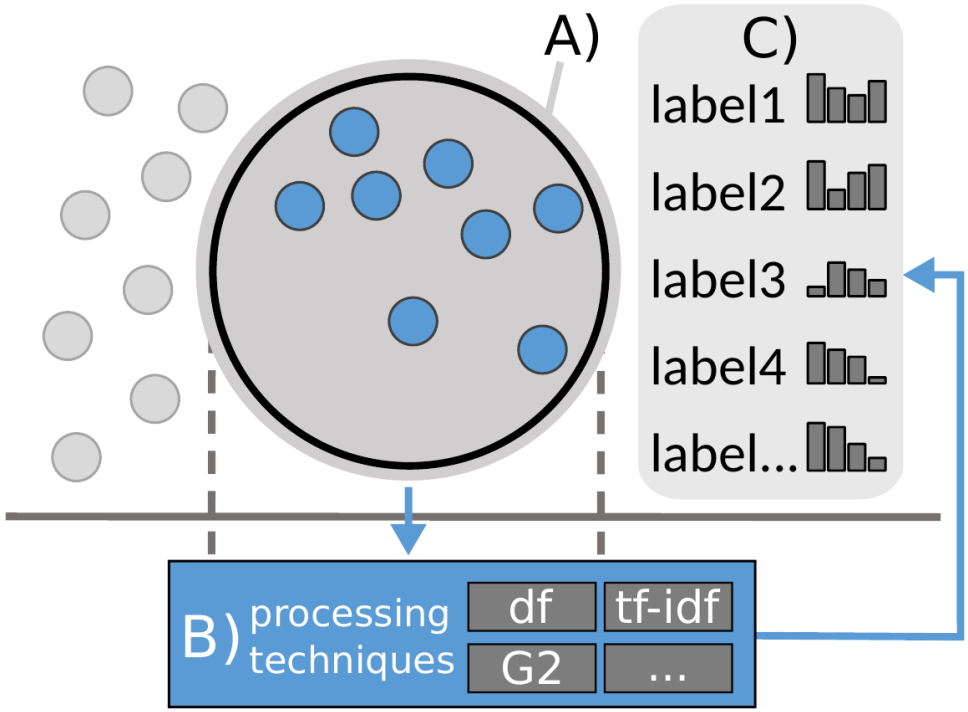
Overview

- Interaction technique

  A: Lens

  B: Parameters

  C: Readable output



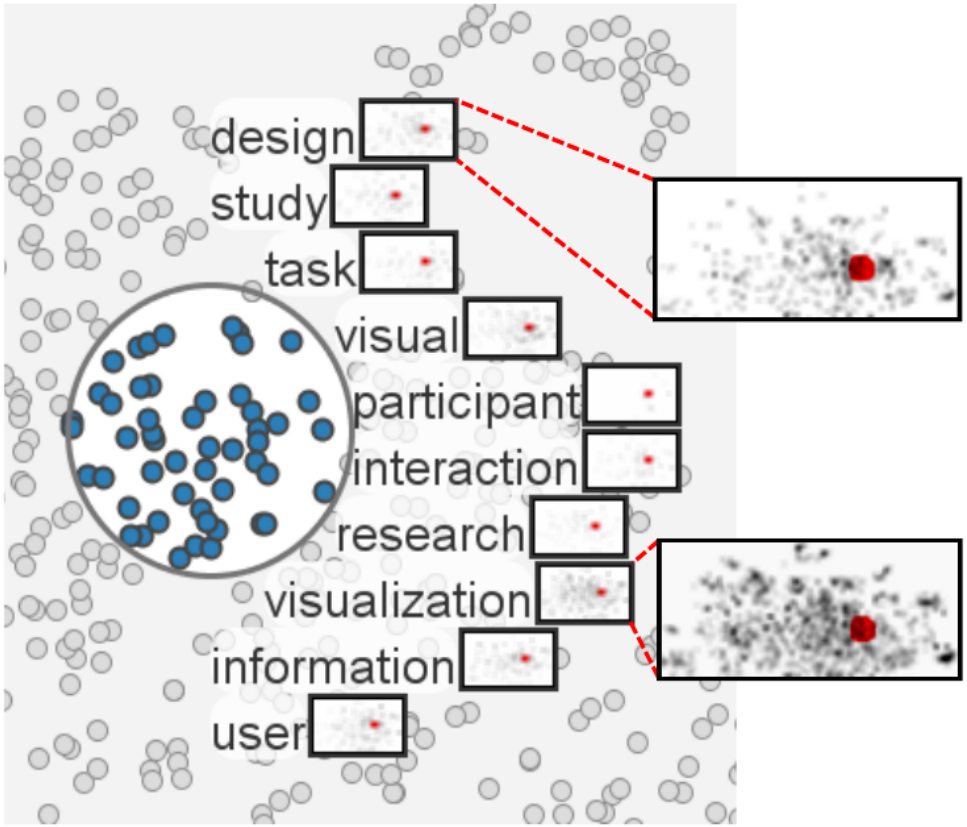
Goals

1. Flexible analysis on arbitrary levels of granularity of the dataset
2. Human-comprehensible visual output to the focused document set
3. Focusing a set of documents and analyse it without causing cluttering or overlaping
4. Global and local scale interaction
5. Scalable to any 2D document dataset

Design I

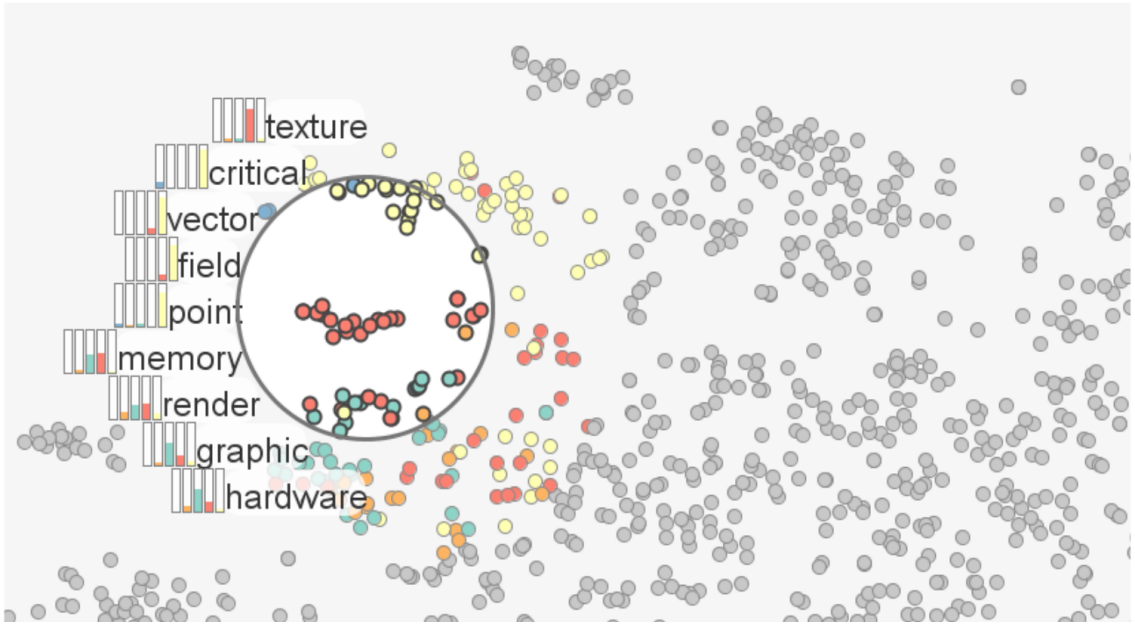
- Placement of Cues

  Goals 1 and 3



Source: Heimerl et al. (2017)

Design II



Source: Heimerl et al. (2017)

- Global Navigation

  Goals 3 and 4
- Local Navigation

  Goal 4

Document Characterization

- Term-based
  - Document frequency (also its inverse alternative)
  - Lemmatization
- Metadata
  - Quantifiable
  - Plottable
- Intuitive and user-comprehensible

  Goals 2 and 5

Document Spacialization

The effectiveness of some of the characterization and navigation methos of DocuCompass depend on the underlying document spatialization.

- Inherent 2D Coordinates
  - Geo-located
- Metadata-based Mapping
- Dimensionality Reduction
  - Information loss
  - Most popular: LSP and tSNE

Table of Contents

Methodology

- User experience
  - Visualization
  - NLP
- User feedback
  - Satisfying

Table of Contents

Conclusions

- Fills a gap between visualization and interaction techniques that provide large scale overview and detailed inspection of text corpora
- A powerful complement to many visual analytics approaches which profit from continuous exploration as part of text foraging tasks
- Flexible and adaptable to different text types and analysis tasks

References I

F. Heimerl, M. John, Q. Han, S. Koch, and T. Ertl, "DocuCompass: Effective exploration of document landscapes," *2016 IEEE Conference on Visual Analytics Science and Technology, VAST 2016 - Proceedings*, pp. 11–20, 2017.

T. Munzner, *Visualization Analysis and Design*. A K Peters/CRC Press, 2014.