InVITe:
Towards Intuitive Visualization of Iterations over Text

Houda Lamqaddam[1, 2], **Jan Aerts**[1, 2]
(1) VDA-lab, ESAT/STADIUS, KU Leuven, Belgium
(2) iMinds Medical IT, KU Leuven, Belgium

‹firstname.lastname›@esat.kuleuven.be

CONTEXT

With InVITe, we are working towards intuitive visualization to support review of iterative modifications on text documents.
In order to accomplish this, we perform simple matching of text snippets between two versions of text, across a range of parameter settings. Next, an overview graphic indicating the effect of parameter space on the output allows the user to select those combinations that are of interest. Finally, such selection will display an alluvial diagram with annotations and covering different resolutions.

With this tool, co-authors can keep an overview of changes made, both structural and local.
A typical use-case for this tool could be the collaborative process between a student and a professor reviewing their work. The student will prepare a draft, which is reviewed by the promoter. The latter has several remarks, which may include small changes as well as substantial restructuring of the text itself. The student implements these suggestions and sends the new version of the text to their promoter. Although current text editing software include text tracking features, these do not allow the promoter to have an overview of the changes brought to the document, causing them instead to traverse the text in its entirety to form a comprehensive understanding of the applied modifications. They also do not present different resolution options to their users.
The InVITe interactive visualization aims at providing a synthetic overview of the evolution of a text through iterations, on a wide range of parameter configurations.

THE TOOL

(1) Parameter space matrix

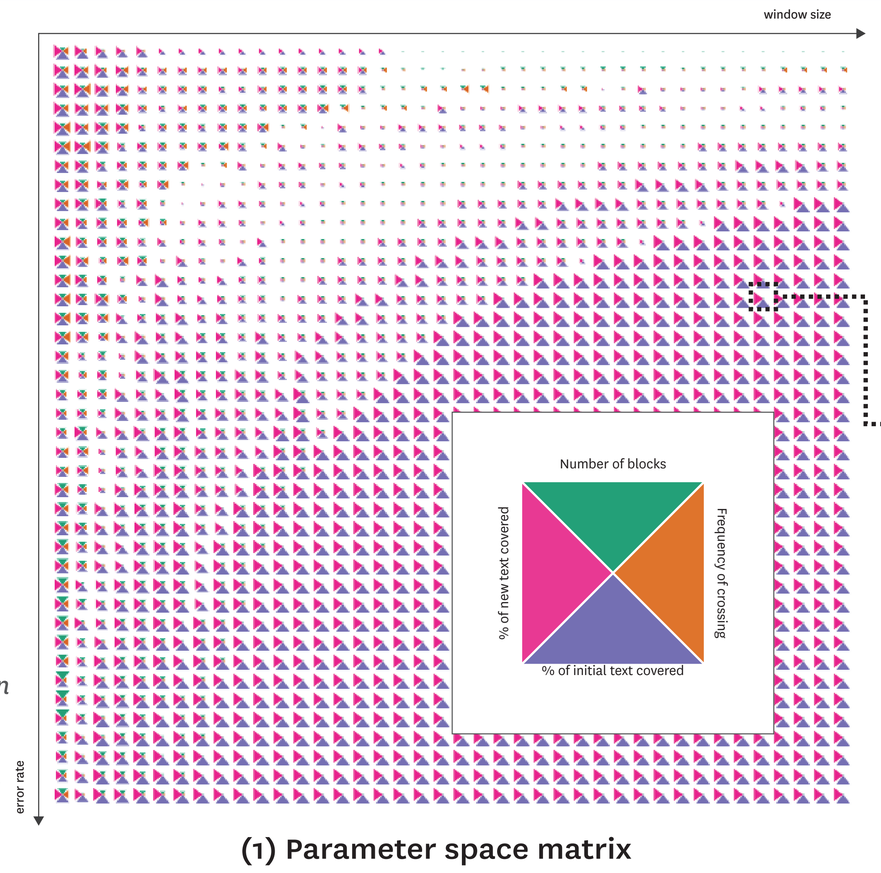
A matrix covering parameter space for window size w and error rate e, and indicating their effect on the resulting blocks. Each cell corresponds to a different instance of the content flow diagram.

Plots in the lower right corner indicate that with high e and w one ends up with a large matches of single blocks; using low e and w (top left) identifies crossings and multiple blocks.

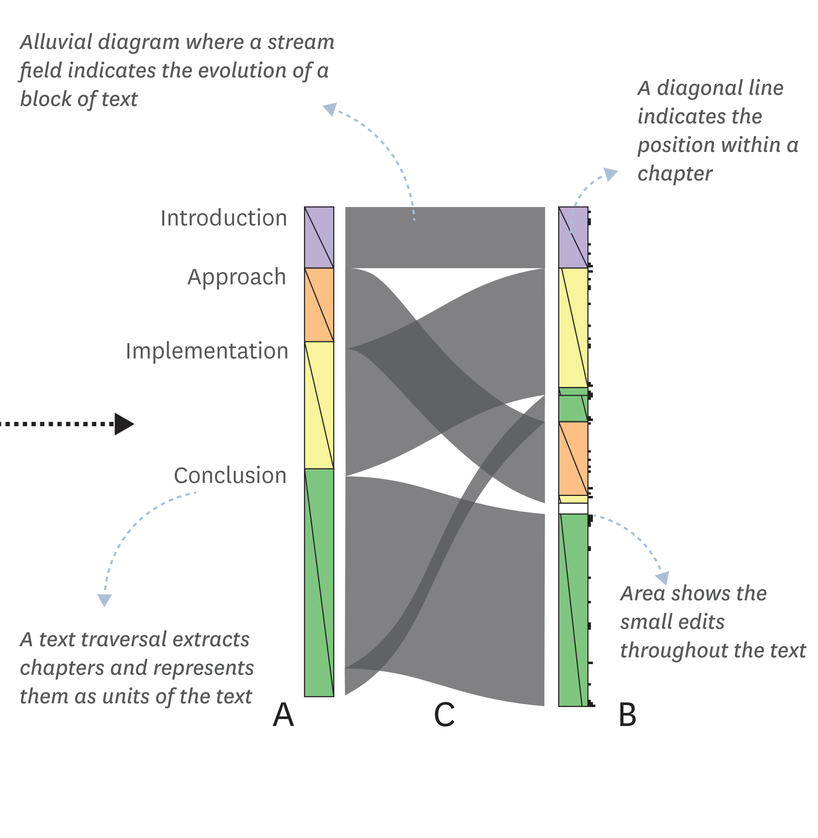
(2) Content flow diagram

An alluvial diagram showing how two versions of text differ from each other, as well as the sensitivity of each section with changes in parameter settings.

Selecting a different square in the matrix will result in a new content flow diagram.



(1) Parameter space matrix



(2) Content flow diagram

APPROACH

Visualisation

As the window size and error margin parameters have a significant impact on the resulting plot, InVITe provides an overview of this parameter space as indicated in Figure 1. This allows the user to choose the **granularity** of the returned visual, corresponding to the task that they want to perform. Selecting a combination of w and e results in the **alluvial diagram** as presented in Figure 2.

In this diagram, **part A** represents the reference text, indicating each chapter in a different colour and including the **section title** if the text is written in **Markdown syntax**. A diagonal line across the box indicates the position within that chapter; an approach regularly used in comparing genomes between species.

Part B shows the new version, in its rearranged state. Chapters of origin and rearrangements within them can be easily identified using the **colour encoding** and **diagonal line**. In addition, the marks on the right indicate stability, showing where breaks in the alluvial diagram would appear if the user were to choose a smaller window size w and/or lower allowed error rate e. Hovering the mouse over a section in either the initial or new version will show the underlying text.

An **alluvial diagram** (**C**) connects the two text versions. Clicking on a streamfield (C) will open a side-by-side view, zoomed into that section, using more stringent parameter settings and therefore splitting the text further into subsections.

Algorithm

The interactive InVITe visualization relies on a simple text analysis approach. The original text is considered the reference and divided into **atoms of a pre-defined window size w.** The algorithm then scans the second text in search of these atoms.

This match does not have to be perfect as we allow an **error rate e** between the reference atom and its match. Using large values for window size w and error e favours **visualization of large structural changes** in the text (e.g. new, deleted or translocated sections), whereas small values favour **small local changes** such as spelling or word choice. Our tests indicate that running 1,600 combinations of the w and e parameters on two versions of a 216-page document takes 21 seconds on a 2,2 GHz Intel Core i7 Mac laptop.

FUTURE WORK

For the next versions of the interactive InVITe visualization, we plan to support multiple texts comparison in order to view the evolution of a text over a multitude of iterations. We also intend to evaluate and improve the currently implemented user interactions.

AKNOWLEDGEMENTS

This work is supported by funding from iMinds Medical IT and VLAIO/SBO project ACCUMULATE.