Boundary Labeling for annotated documents

Jakob Klinger
Matr.: 1125755
Curriculum: 033 534
e1125755@student.tuwien.ac.at

July 24, 2017

1 Introduction

Annotating a document is usually solved by adding footnotes or figures in an appropriate position and adding a simple reference in the text, leaving the reader to find the referenced content by themselves. Sometimes however, if a more obvious connection between the text and the referenced content is required, the reference is visibly connected to the text by drawing a line between them.

In this paper, we will look at ways to use Boundary Labeling, which means that all annotations will be placed outside of the text they are referencing. (See also [1]) While there are many papers discussing Boundary Labeling in general, only very few exist that apply this concept to written text. As a result, and since it is very easy to accidentally obscure parts of the text, this approach isn't used very often, and tends to use simplistic algorithms which produce mediocre results.

2 Boundary Labeling

Boundary Labeling can be explained very quickly - connect a Point of Interest with a box containing further information which is located outside of the main area. To easily reference important concepts, some additional Terminology will be introduced (See Fig. 1 for a visual explanation)

- **Label:** The additional information, usually represented as a box containing the additional information. Will also be referred to as **"Annotation"** in this paper.
- **Site:** The point, object or word that is labeled - without it, there wouldn't be any labeling necessary.
- **Leader:** The line connecting the site to the label - depending on its shape, it can be further classified into several subgroups which can be freely combined to form new leader types:
  - *S-Leader:* This Leader connects site and label in a straight line.
  - *O-Leader:* This leader runs orthogonally to the border between text and label area.
  - *P-Leader:* This Leader runs parallel to the border between text and label area. It must be combined with other leader styles, as it won't reach the label area otherwise.

  For example, the leader from Fig. 1 would be classified as an opo-Leader.
- **Port:** The location where the leader connects to the label. It may be restricted pre-determined locations, like only at the corners.

2.1 Boundary Labeling in Documents

As with other labeling techniques, Boundary Labeling has guidelines on how to create an optimal labeling - the leaders should be as direct as possible, no important information should be obscured, and it should be easily discernable which Label belongs to which site. These three often stand in conflict with one another, especially when labeling documents, as the text usually is very dense and leaves little space for leaders in between, yet one shouldn't allow them to pass through the text, as this makes the text harder to read.

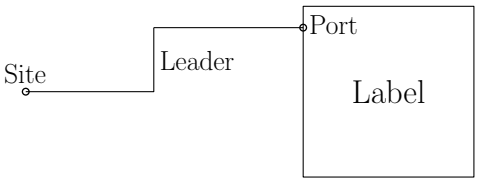

Figure 1: Illustrated guide to the labeling terminology

3 Previous Attempts

While there haven't been many papers written about this topic,
they do offer interesting contributions. For example, the paper about the Luatodonotes-Package[2] illustrates some of the different leader types available, and came to the conclusion that leaders without bends are easier to follow. However, most solutions proposed in that paper didn't care whether a leader overlapped with text or not, which resulted in a decrease in readability.

The paper by Loose[3] on the other hand is based around only using the free space between the text, which produces slightly longer leaders, and forces curves. In short, the Leaders can be described as a sequence of O- and P-segments, which continue until the label's port is reached.

References

[1] Michael A. Bekos, Michael Kaufmann, Antonios Symvonis, and Alexander Wolff. Boundary labeling: Models and efficient algorithms for rectangular maps. *Computational Geometry*, 36(3):215 – 236, 2007.

[2] Philipp Kindermann, Fabian Lipp, and Alexander Wolff. *Luatodonotes: Boundary Labeling for Annotations in Texts*, pages 76–88. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[3] Amrei Loose. Annotation von texten unter berücksichtigung von textzwischenräumen. Master's thesis, Karlsruhe Institute of Technology, 2015.