

Characterizing Annotations for Visual Analytics on Clinical Data

Dissertation

A thesis submitted for the degree of

Doktor-Ingenieur (Dr.-Ing.)

Faculty of Computer Science and Electrical Engineering
University of Rostock



Christoph Schmidt
born on July 9th 1978 in Rostock
resident in Schwerin

Rostock, 17.04.2024

Kurzfassung

Annotationen sind nach dem Verständnis in dieser Arbeit Informationen die zusätzlich in ein Visual Analytics System eingebracht werden. In der Literatur gibt es hierzu bereits zahlreiche Beispiele. Viele dieser Beispiele zeigen die konkrete Umsetzung von Annotationen, die in einem speziellen Anwendungsfall unterstützend wirken. Eine systematische Analyse der Annotationscharakteristiken hingegen ist nach wie vor Gegenstand der Forschung und eine große Herausforderung. Mit Hilfe einer solchen Analyse wäre es möglich, angemessene Charakteristiken sowie ein passendes Design für Annotationen systematisch zu entwickeln, also ein *Annotation-sproblem* zu lösen. Wir adressieren die systematische Lösung von Annotationsproblemen, indem wir zunächst die Charakteristiken vorhandener Annotationslösungen herausarbeiten und anhand grundlegender Fragen zu Annotationscharakteristiken in einer morphologischen Box strukturieren. Darüber hinaus entwickeln wir ein Modell, welches die morphologisch Box für die weitere Analyse zugrunde legt. Unter Heranziehung der Anforderungen aus dem Anwendungsfall sowie aus der Visuellen Datenanalyse können mit Hilfe des Modells angemessene Annotationscharakteristiken ermittelt werden. Hierdurch müssen die Annotationsmöglichkeiten im konkreten Anwendungsfall nicht mehr aufwändig aus allen Möglichkeiten herauskristallisiert werden, sondern werden mit Hilfe des Modells systematisch entwickelt. Im weiteren Verlauf zeigen wir, wie man ein passendes Design für die Annotationen mit den angemessenen Annotationscharakteristiken entwickeln kann. Dieses Design steht dabei in direkter Beziehung zur Visualisierung der Originaldaten und erfüllt die Anforderungen sowohl der Originaldatenvisualisierung als auch die Anforderungen der Annotationsvisualisierung. Um die Anwendbarkeit nachzuweisen, implementieren wir das Annotationsdesign in ein bestehendes Visual Analyticsystem. Mehrere Sitzungen zur Erlangung von User Feedback und die testweise Anwendung des Modells auf epidemiologische Daten zeigen, dass diese Annotationsermittlung zu einer besseren Datenvorbereitung, Datenbereinigung und Exploration von heterogenen klinischen Daten beitragen kann.

Abstract

Annotations in this work are understood as supplementary information integrated into a visual analytics system. Various examples in literature show their usefulness and application. However, many of these examples contain a specific annotation solution to support the analysis of a particular use case. A systematic approach to analyze and develop the annotation characteristics is still ongoing research and difficult to address. Such an analysis could help to systematically characterize and design annotations in accordance with the needs of a particular use case, in short: to solve an *annotation problem*. We address the development of a systematic solution for the annotation problem by conducting a literature survey on existing annotations and extracting the annotation characteristics. We structure these characteristics based on basic questions on annotations and sort them into a morphological box. Furthermore, we develop an annotation characteristics model, which takes the morphological box as a base. The model additionally takes the requirements from both the use case and visual analytics into account and derives suitable annotation characteristics for this particular use case. This helps domain experts and visual analytics experts to systematically develop suitable annotations. Additionally, we show how a fitting design for the annotations can be developed for the derived suitable annotations. This design is aligned with the original data visualization and fulfills both the requirements from the original visualization and annotation visualization. To show the usability we implement the annotations into a visual analytics system for heterogeneous clinical data. Several user feedback sessions and a brief approach to apply our model to epidemiological data show that our approach can help to improve data preprocessing, data cleansing and data exploration for clinical data.