

Abstract

Placing labels that contain text or images is a common task in information visualization. Labels convey information about objects in graphical displays like graphs, networks, diagrams, or cartographic maps. Typically, each object (or *feature*) that has to be labeled allows a number of positions where the corresponding label can be placed. However, each of these label candidates could intersect with label candidates of other features. This gives rise to the following combinatorial problem, the *general label-placement decision problem*. Given a set of features, each with a set of label candidates, can we assign each feature a label from its candidate set such that no two labels intersect? In other words, is there a *complete labeling* for the set of features?

Unfortunately, one cannot expect to find an efficient method for answering this question. Therefore most previous work has focused on developing heuristics and approximation algorithms for various optimization problems related to label placement, such as finding a placement for a subset of the features of maximum cardinality that has a complete labeling. This problem is referred to as the *label-number maximization problem*.

In this thesis we approach label placement from two sides. On the one hand we develop a general framework for label-placement problems. This framework extends the classical constraint-satisfaction framework such that the label-number maximization problem can be formulated and algorithms can be developed that reduce the size of the search space for an optimal solution considerably. We give such an algorithm and a simple, but very effective heuristic based on this algorithm.

On the other hand we investigate special cases of the label-placement problem, which is generally divided into point, line, and area labeling. First we consider the point-labeling problem. We apply our framework to labeling points with axis-parallel rectangles, which can, for example, represent the bounding boxes of textual labels. We allow the classical four label positions per point, namely those where a corner of the rectangle coincides with the point. We compare our method to five other point-labeling algorithms experimentally. Then we investigate point labeling with an infinite number of label candidates per point. We show that it is NP-hard to decide whether a set of points can be labeled with unit squares or with unit disks if we require that each label touches its point. Nevertheless we give efficient approximation algorithms for optimization versions of both problems. More specifically we give a polynomial-time

approximation scheme for maximizing the number of axis-parallel rectangular labels of common height, while we show that one cannot expect to find such a scheme for maximizing the size of uniform circular labels.

For labeling polygonal chains like rivers or railway lines on maps, we study the cartographers' requirements and put them into two categories; hard and soft constraints. Then we give an efficient algorithm that guarantees to satisfy all hard constraints. In addition we show how to optimize the soft constraints. The method we suggest is the first that simultaneously fulfills both of the following two requirements: it allows curved labels and its runtime is at most quadratic in the number of points on the given polyline.

Apart from analyzing asymptotical runtime behavior and storage requirements of our algorithms, we implemented most of them and studied their performance on synthetic as well as real-world data. The last chapter of this thesis is devoted to generic programming, a method for abstracting from concrete data representation, that we found very helpful for keeping our geometric algorithms flexible.

Zusammenfassung

Bei der Visualisierung von Information auf Landkarten, in Graphen oder Diagrammen spielt die Beschriftung von Gestaltungselementen wie Punkten, Linienzügen oder Kanten eine große Rolle. So schätzt man, dass bei der manuellen Erstellung einer Landkarte ungefähr die Hälfte der Zeit für die Beschriftung benötigt wird. Dieser Umstand erklärt das Interesse daran, den Prozess des Beschriftens möglichst weitgehend zu automatisieren. Das Beschriftungsproblem lässt sich ganz allgemein als kombinatorisches Entscheidungsproblem ausdrücken: Gegeben eine Menge von zu beschriftenden Elementen einer Grafik und zu jedem Element eine Menge von Beschriftungskandidaten, ist es möglich, jedem Element einen Kandidaten aus der betreffenden Menge zuzuordnen, ohne dass sich zwei der gewählten Kandidaten schneiden?

Leider muss man davon ausgehen, dass dieses Problem im allgemeinen nicht effizient zu entscheiden ist. Daher haben fast alle bisherigen Arbeiten zu diesem Thema heuristische oder approximative Verfahren für entsprechende Optimierungsprobleme vorgeschlagen, etwa um eine Beschriftung einer möglichst großen Teilmenge der Grafikelemente zu finden. Dieses Problem bezeichnet man als das Anzahlmaximierungsproblem.

In meiner Dissertation näherte ich mich dem Beschriftungsproblem von zwei Seiten. Einerseits stelle ich ein theoretisches Gerüst auf, mit dessen Hilfe man das Anzahlmaximierungsproblem für endliche Kandidatenmengen gut formulieren und effiziente Algorithmen angeben kann, die den Suchraum für eine optimale Lösung erheblich verkleinern. Dieses Gerüst verallgemeinert das klassische Constraint-Satisfaction-Problem. Ich gebe einen solchen Algorithmus sowie eine einfache Heuristik an, die auf diesen Algorithmus aufbaut und in der Praxis sehr gute Resultate liefert.

Andererseits beschäftige ich mich mit Spezialfällen des Beschriftungsproblems. Ich habe mich zuerst mit der Beschriftung von Punkten mit achsenparallelen Rechtecken befasst, wobei für jeden Punkt die vier klassischen Positionen zugelassen wurden, also die, bei denen eine Ecke der Beschriftung den betreffenden Punkt berühren muss. Auf dieses Spezialproblem habe ich den erwähnten allgemeinen Algorithmus angewandt und mit anderen Verfahren experimentell verglichen. Dann habe ich mich mit Beschriftungsmodellen beschäftigt, in denen eine unendliche Anzahl von Beschriftungskandidaten zugelassen wird. Ich konnte zeigen, dass man nicht erwarten kann, obiges Entscheidungsproblem für quadratische oder kreisförmige Beschriftungen effizient zu lösen, falls jede Be-

schriftung ihren Punkt berühren muss. Trotzdem habe ich effiziente Algorithmen für Optimierungsversionen beider Probleme gefunden. Für achsenparallele rechteckige Beschriftungen gleicher Höhe stelle ich ein Approximationsschema für das Anzahlmaximierungsproblem vor. Andererseits zeige ich, dass man nicht erwarten kann, ein solches Schema für die Maximierung der Größe kreisförmiger Beschriftungen zu finden.

Ausser der Beschriftung von Punkten untersuche ich das Problem, wie man Linienzüge, also Flüsse oder Straßen, qualitativ hochwertig beschriften kann. Dies ist im Gegensatz zur Punktbeschriftung, wo die Zielfunktion meist klar ist, eher ein Modellierungsproblem, bei dem man die Anforderungen der Kartographen erst herausfinden muss. Ich habe diese dann in harte und weiche Anforderungen eingeteilt und einen effizienten Algorithmus vorgeschlagen, der garantiert, die harten Anforderungen zu erfüllen, und der innerhalb eines Teils des Suchraums die weichen Anforderungen soweit wie möglich optimiert.

Um die Praxisrelevanz meiner Untersuchungen zu unterstreichen, sind die meisten angegebenen Algorithmen implementiert worden. Dabei bin ich auf das schon bekannte Paradigma des generischen Programmierens gestoßen, das es ermöglicht, Programme sehr allgemein und flexibel zu halten. Ich zeige, wie man damit erfolgreich geometrische Algorithmen entwirft.