

# Area segmentation algorithms - A Geoinformatic approach to Geomarketing strategies

Sabine Schmidt  
Master of Geoinformatics  
Matriculation register  
In cooperation with microm

1th of October 2015

First supervisor:  
Dr. Torsten Prinz (WWU)  
Second supervisor:  
Dipl. Ing. Sven Reifegerste (microm)

# Contents

<b>Abstract</b>	<b>4</b>
<b>Acknowledgement</b>	<b>5</b>
<b>1 Introduction</b>	<b>6</b>
1.1 Definition and aim of Geomarketing	6
1.2 microm Micromarketing-Systems and Consult GmbH	8
1.3 Motivation and Research Question	10
1.4 Methods	10
1.5 Outline	10
<b>2 Related Work</b>	<b>11</b>
2.1 KIT - Institute of Operations Research: discrete optimization and logistic	11
2.2 Easy Map District Manager	14
<b>3 Application of area segmentation</b>	<b>17</b>
3.1 Introduction	17
3.2 Parameters	17
3.3 previous researches	17
3.4 Use cases	17
3.4.1 Sales Districting	17
3.4.2 Political Districting	17
3.4.3 Greenfieldanalysis	18
3.4.4 Whitespotanalysis	18
<b>4 Area segmentation approaches and their implementation</b>	<b>19</b>
4.1 Segmentations considering homogenous distribution	19
4.1.0.1 Considering just Criteria	19
4.1.0.2 Considering average criteria value in proportion to number of locations	19
4.2 Segmentations considering distance	19
4.2.0.3 Considering just Distance	19
4.3 Segmentations considering criteria and distance	19
4.3.0.4 Criteria and Distance: from inside to outside: Small- estCritGetsNearest	19
4.3.0.5 Criteria and Distance: from inside to outside: Small- estCritGetsTrueNearest	19

4.3.0.6	Criteria and Distance: from outside to inside + inside to outside . . . . .	19
4.3.0.7	sum of criteria divided by number locations + Distance . . . . .	19
4.3.0.8	Distance + criteria . . . . .	19
<b>5</b>	<b>Problems of the implemented approaches.....</b>	<b>20</b>
5.1	Performance .....	20
5.2	Requirements from the field of Geomarketing.....	20
5.2.1	Incoherent areas.....	20
5.2.2	Inhomogeneous distribution .....	20
5.2.3	Rearrangement / Infinite loops during rearrangement .....	20
5.2.4	Need of threshold Values .....	20
<b>6</b>	<b>Comparison of implemented approaches.....</b>	<b>21</b>
6.1	Performance .....	21
6.2	Problems .....	21
6.3	Requirements.....	21
6.4	Conclusion .....	21
<b>7</b>	<b>Application of algorithm XYZ to Geomarketing analysis.....</b>	<b>22</b>
7.1	Optimization of Areas .....	22
7.1.1	Conditions and Aims .....	22
7.1.2	Approach of algorithm .....	22
7.2	Greenfieldanalysis.....	22
7.2.1	Conditions and Aims .....	22
7.2.2	Approach of algorithm .....	22
7.3	Whitespotanalysis.....	22
7.3.1	Conditions and Aims .....	22
7.3.2	Approach of algorithm .....	22
<b>8</b>	<b>Realworld scenario: integretaded Algorithm to the mapChart Manager .....</b>	<b>23</b>
<b>9</b>	<b>Evaluation.....</b>	<b>24</b>
<b>10</b>	<b>Discussion and Perspective .....</b>	<b>25</b>
10.1	Summary .....	25
10.2	Limitations.....	25
10.3	Comparison to related work.....	25
10.4	Perspective.....	25

# Declaration

Abstract

Abstract

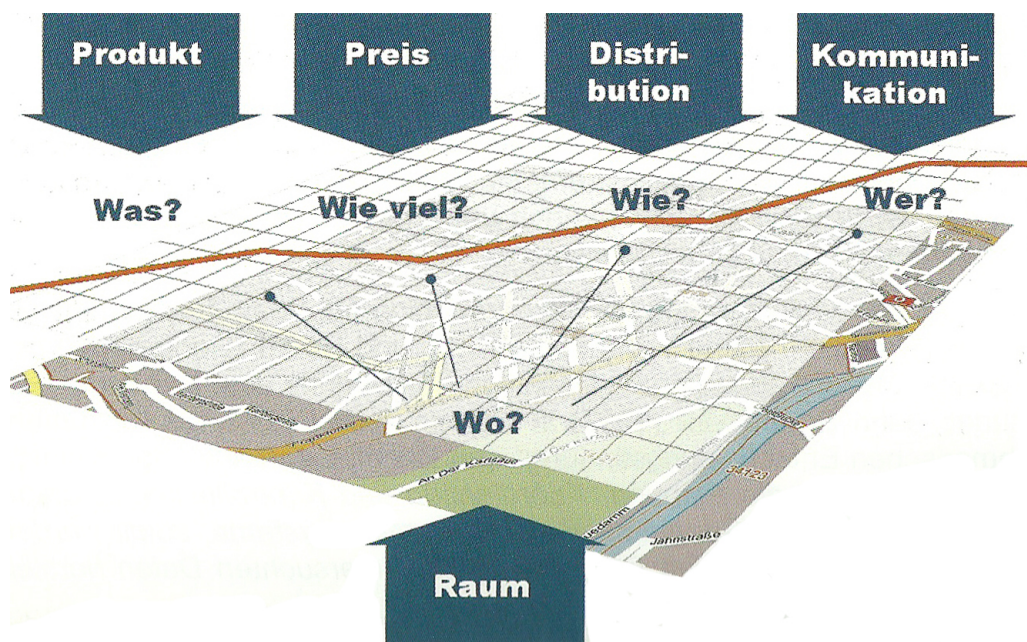
# Acknowledgement

# 1 Introduction

## 1.1 Definition and aim of Geomarketing

The term of Geomarketing has established more and more in the field of marketing within the last years. A first approximation to the notion of Geomarketing was done in 1995 by Frühling and Steingrube. They had explained that Geomarketing is just a genus for several instruments within the field of marketing. This shows that Geomarketing is no methodology but rather a discipline. Although some definitions had occurred in the 90s, the first use of Geomarketing analysis went back to the 50s. Already 1952 the first map showing the purchasing power of Germany was published. In 1982 several companies were founded, which have offered tools and possibilities for their costumers to practise geographic analysis so that these researches got easier more and more. Consequently the comprehensive application of Geomarketing was born [3]. Within the early 90s Geomarketing described approaches and fundamentals of geographic analysis and the governance of marketing and distribution. The central idea of Geomarketing is that marketing composed of price, product, distribution and communication will be complemented by space. Consequently operating numbers can change dependent on the spatial location by spatial phenomena of productions and logistics.

**Figure 1**  
Generic  
consideration of  
Geomarketing  
aspects  
complemented by  
space [3]

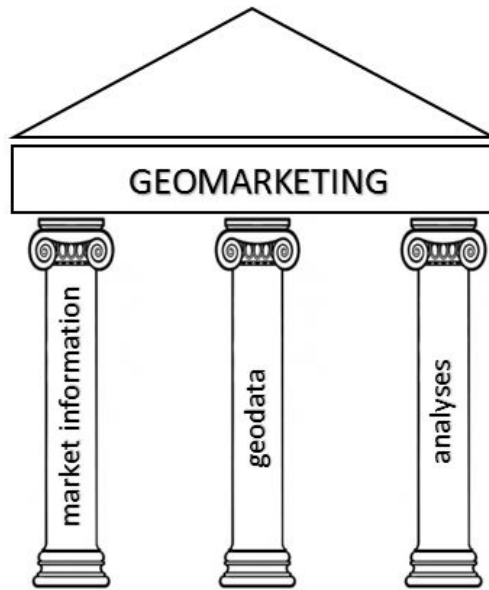


Although the central idea of Geomarketing is known the definition of Geomarketing leads to some problems so that often incorrect explanations are written down. It is often readable that Geomarketing is a spatial analysis of the market using a

geographic information system. But by particularly consideration it is recognizable that this statement is not correct because a geographic information system is just a tool which supports Geomarketing analysis to make them easier. GmbH tried to find a definition of Geomarketing which fits best compared to other definitions. They have defined that Geomarketing analyses current as well as potential markets considering spatial structures to make the planing of product sales more effective and to control them quantifiable. That means that all available information about the market are connected to a spatial reference system to make dependences, potentials and other properties visible. The application of Geomarketing analysis have their origins in the minimization of entrepreneurial risks by making the market more transparent and acting purposeful. In the course of this Geomarketing has established as a sub-discipline of the field of marketing. During the application of Geomarketing analysis several benefits can be achieved. By knowing potential costumers and competitors the marketing and distribution of products from a company can be done more dedicated so that efficiency enhancement and cost reduction can be caused. Additionally advantages to the competitors are achieved by the analyses. Furthermore inquests of the market may be helpful during the planing of new locations to determine a site with a high potential so that the risk of a malinvestment can be minimized. It is recognizable that as higher as the number of costumers is, none the worse the benefit of Geomarketing analyses are. During the surveys of the market several principles are utilized. One of them is the spatial factor of the market. Using spatial data of costumers, locations, branch offices etc. important dependences can be visualized. The spatial data are almost given using addresses. Additionally a spatial heterogeneity can be recognized during analyses that means that the market differentiates in space. in conclusion a third principle is generated by that fact. It describes the spatial segmentation of the market. It is recognizable that a subdivision into homogeneous market segments is possible. Consumers in the western part and the eastern part of Germany differentiate to each other, but an identification of costumers with similar affectations within a small range of space is possible. From this it follows the neighbourhood principle which explains that neighboured customers have a similar behaviour considering marketing aspects as product purchase. This fact has two reasons. On the one hand costumers with an analogical lifestyle live in the same space and consequently show common characters in costumer behaviour. On the other hand these people share the same infrastructure which takes influences to their purchasing habits as well. Additionally the distance to the location of a company affects the costumers in their decisions going to it or not. Geomarketing is based on three stacks: information of the market, geodata and analyses.



**Figure 2**  
*Basic stacks of  
 Geomarketing*



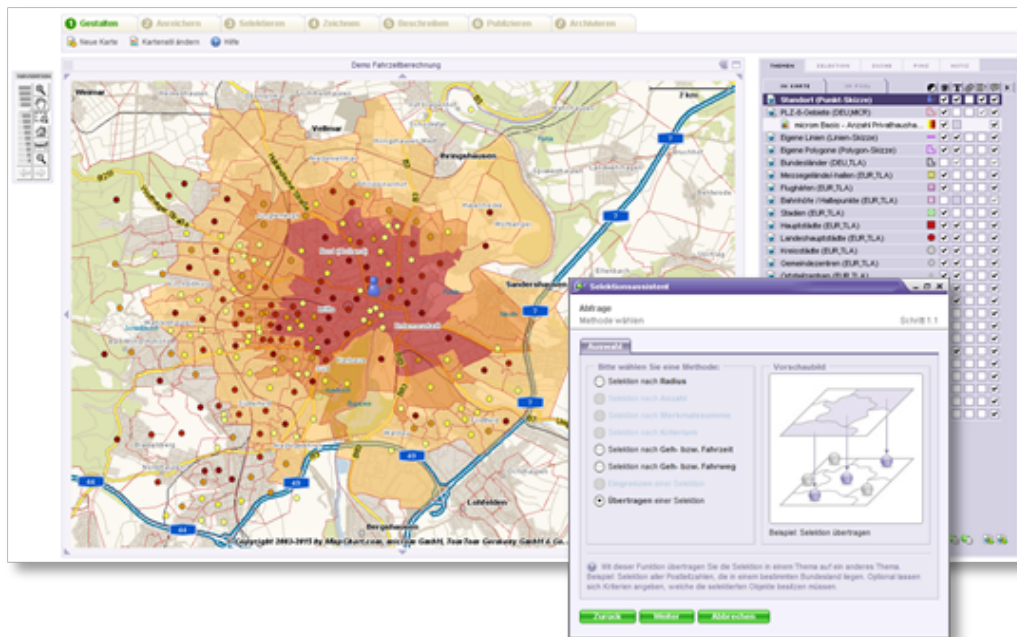
Market information are qualitative facts about costumers, competitors etc. within a regional (economic) zone. The data contain information about their socio-demographic, psychosocial, economic and consumption properties like income, product affinity, gender and household size [? ]. Geodata are information with a spatial reference like addresses, sales areas, locations and catchment areas. The boundaries of these regions may be administrative borders like federal states or townships as well as street sections, coordinates of houses and individual created areas per example a subdivision of postal code areas. By connecting market information with the help of geocoding to these geodata analyses are possible to obtain important knowledge about the market which are helpful to support companies in their marketing decisions. These facts show that Geomarketing is an instrument for analysing, planning, checking and controlling the market. In the meantime Geomarketing is grown up to one of the most important approaches within the field of Geomarketing to support companies while the accomplishment of their strategies. Consequently it is getting more and more essential to have systems providing functions and tools which making these analyses easier and more efficient.

## 1.2 microm Micromarketing-Systems and Consult GmbH

Microm Micromarketing-Systems and Consult GmbH is a company in Germany which provide Geomarketing analyses to their costumers. It was founded in 1992 and since 1997 it is a subsidiary of the Creditreform. During the last decades microm grew up to one of the biggest providers of micromarketing and Geomarketing within Germany. It offers possibilities and tools to do anaylses of Geomarketing data. With the help of projects company owners can give their data to microm so

that the analyses will be done by them. This approach offers the advantage that the company can use all the knowledge which is provided by the employees of microm to control further steps of their marketing decisions. Besides that procedure microm offers additionally a web tool to their customers so that they can do the analyses by their self. The software is called mapChart Manager and is accessible with the help of a web browser like Firefox or Google Chrome.

**Figure 3**  
*Screenshot of the  
mapChart Manager*



The advantage of a web tool like the mapChart Manager is that the users can have access to their data and maps from all over the world. Consequently sharing results and working independently from a computer and location makes the application of Geomarketing analyses more easily. The mapChart manager offers such functionality like the import of data, geocoding of addresses and do analyses like catchment areas and driving distance zones. To do all that analyses a lot of data is indispensable like routing networks or information about the behaviour of potential customers. All these data are offered by the microm so that their customers can buy the information they need. Doing so microm profits from their affiliation to the Creditreform, which collects customer data from different resources among other things. As a subsidiary of the Creditreform the microm can sell all the information which the Creditreform have been collected.

## 1 Introduction

### 1.3 Motivation and Research Question

### 1.4 Methods

blabla

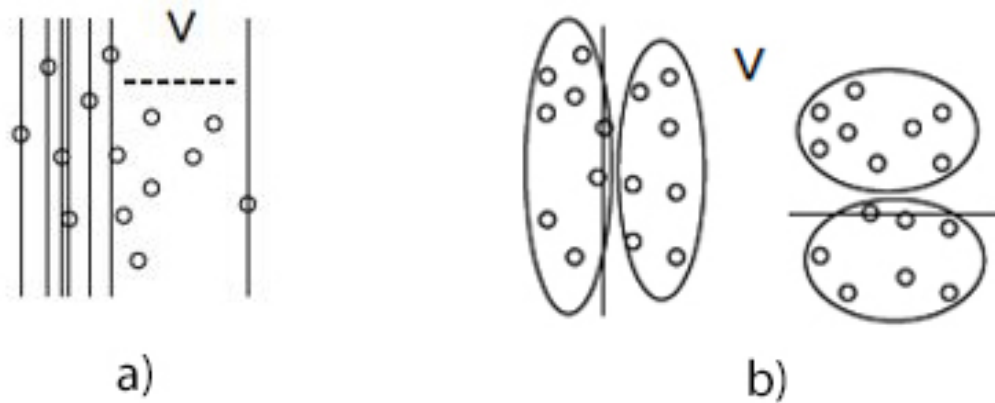
### 1.5 Outline

## 2 Related Work

### 2.1 KIT - Institute of Operations Research: discrete optimization and logistic

The Karlsruhe Institute of Technology (KIT) was founded in 2006 by the amalgamation of the university of Karlsruhe and the research center Karlsruhe. Together both institutes form one of the biggest research and education constitution in the world which concentrates on selected research areas. On the Institute of Operations Research the staff members delve into the field of discrete optimization and logistic which includes districting, too. In progress of studies they have been implemented a library called Lizard which includes an algorithm for the solution of area districting problems. The library can be downloaded for free, so an execution on given example data is possible. In addition to this, it is implemented into a Geographic Information System in the web browser. The algorithm of Lizard is based on the so called "Recursive Partitioning Algorithm" which was designed by Kalcsics et al.. The algorithm is a geometrically approach which divides the problem into sub problems to align territories that are balanced with respect to an activity value. the partition is done by placing a line into the dataset. Consequently a left problem (data left from the line) and a right problem (data right from the line) are created. Within the next steps additional lines are placed into the sub problems to divide them again into smaller ones. The number of sub problems depends on the number of distracted areas which should be created. The lines can have different directions which will be considered. These are all considered so that different partitions of sub problems exists. The sub problems are described by a binary tree whose root element complies the problem that should be solved. The different partition for the subdivisions are ranked by a heuristic measure for balance and compactness so that the partition with the best measure value can be chosen as result.

**Figure 4**  
*a) Possible partitions using vertical lines,*  
*b) Two possible partitions with different compactness [5]*



The steps of the partitions are presentable within Lizard. The download package of the algorithm already includes such example files which can be executed. The examples contains either point coordinates which are connected to an activity or line segments to visualize street segments. These are linked to an activity value too. It is recognizable that no given centres exists. Consequently the area distraction is not done considering given locations like it will be done during the approaches of that thesis. Before the algorithm in Lizard can be started some parameters need to be chosen. There exist input options for:

- Number of areas: This number defines how much parts should be created.
- Number of line directions: These are necessary to define the position of the line which will be placed into the problem to divide into smaller parts.
- Balance Tolerance: The value determines the range which needs to be satisfied comparing the activity of the areas.
- Weight Balance: The value needs to be between 0 and 1 to determine the importance of a balanced activity of the areas. In case of 1 balance is most important, in case of 0 the balance is not important. Instead of this the compactness is most important.
- Compactness Measure: There exist different options how the compactness value is calculated. Some of the options call `ConvexHullIntersection`, `MaximumDistance` and `WeightedPairwiseDistance`. `ConvexHullIntersection` par example uses the circumference of the convex hulls of all district during the calculation of the compactness [11]. Consequently every mentioned option defines another approach of the calculation of the compactness rate.
- Bisecting Partition: With the help of this option the geometric object which divides the problem is selectable. The predefined option are lines how it is defined within the "Recursive Partitioning Algorithm". Additionally `FlexZone` and a combination of both can be chosen.

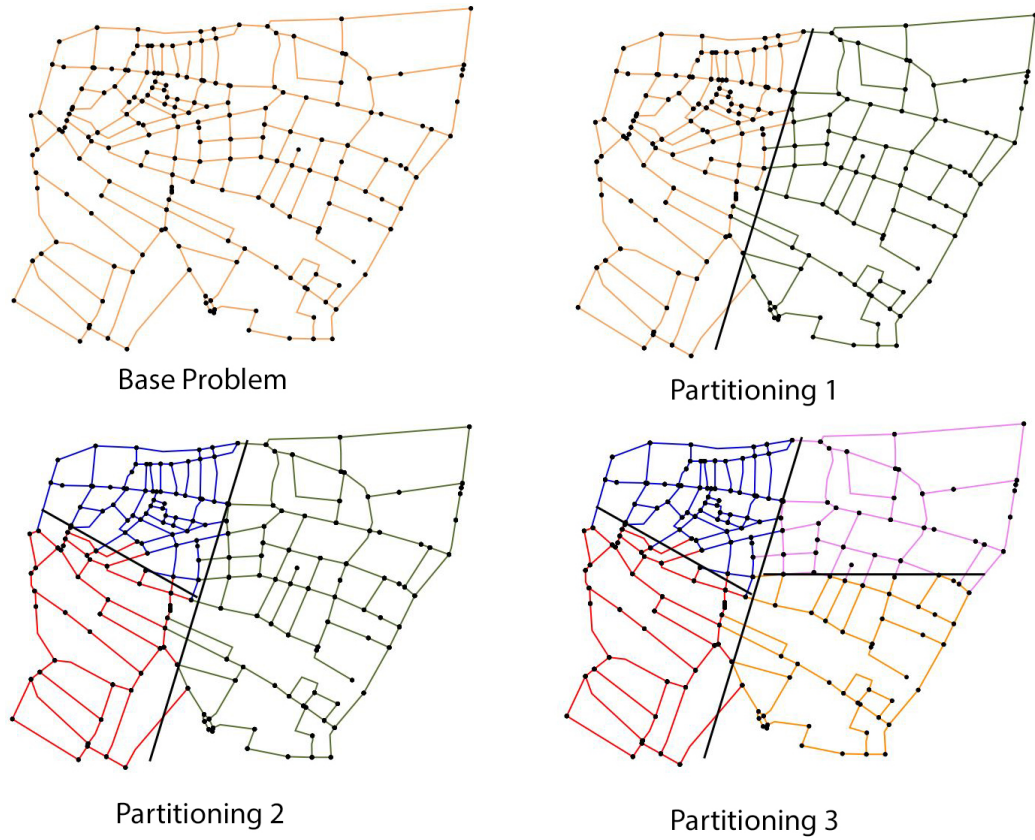
The visualisation of the result of the districting problem can be done step by step or showing just the final result. Showing the results in partial processes represents the approach of the algorithm dividing the problem into sub problems. The calculation of the tree is not recognizable, but the leaf with the best measuring value and thus the best calculation will be shown. Additionally information about other partition solutions are shown. In the following figure the algorithm is visualized for a districting problem where street segments should be divided into four balanced

areas. The feasible balance tolerance is 0.1, the number of search directions are eight, the weight for the balance is 0.5 and the bisecting partition is done by a line. The results of the activity value for every sub regions are:

**Table 1**  
*Activity values of  
resulting areas*

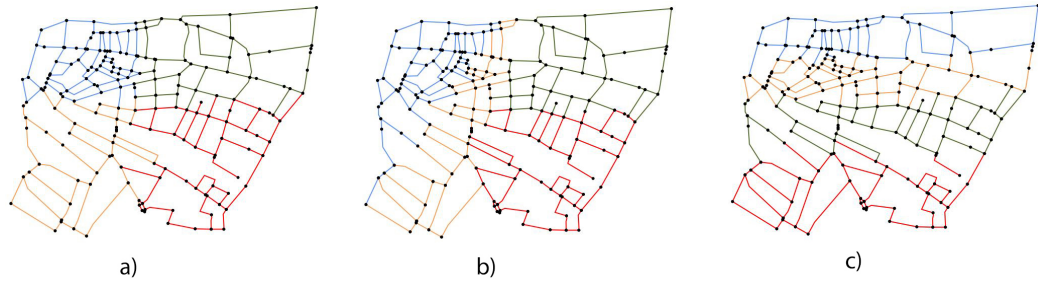
Blue area	Orange area	Green area	Red area
12224	12208	12216	12210

**Figure 5**  
*Visualization of  
area districting in  
Lizard*



Dependent on different settings concerning the available options during the calculations different results can be achieved. Some examples are shown in the following figure. All calculations are done on the same data set as in the figure before. Just the settings where changed. In figure a) the importance of balance balance was set to 0, instead the compactness of the resulting areas should be most important. In figure b) the options is the other way around as in a), consequently the compactness is less important and the balance is most important. Figure c) shows the result of the districting problem if just one search directions are allowed for the lines which were put into the dataset.

**Figure 6**  
*Results of the  
 problem using  
 different options*



## 2.2 Easy Map District Manager

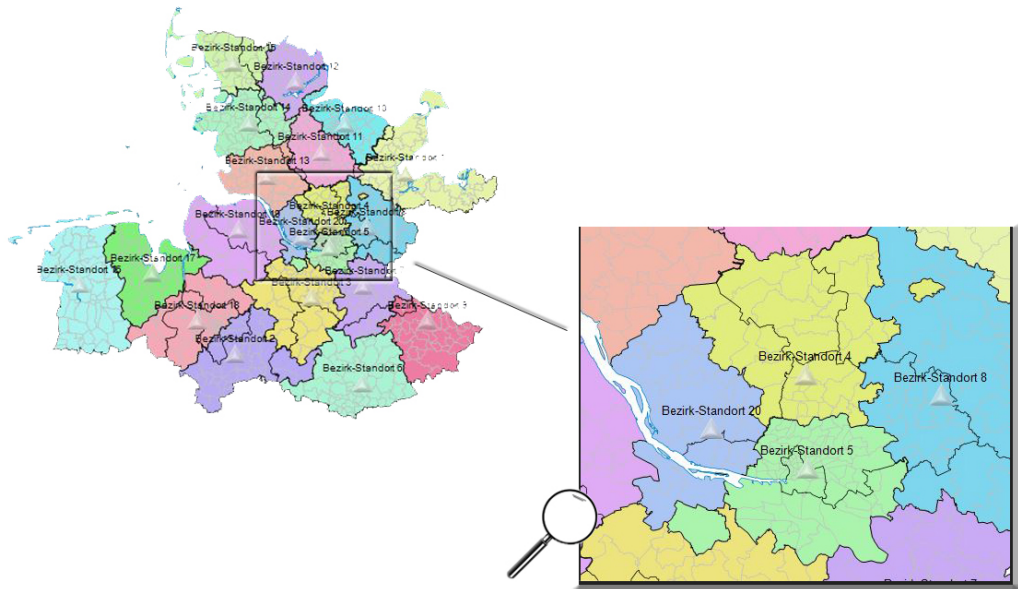
The Lutum + Tappert DV-Beratung GmbH is one of the biggest companies in Germany which provide tools, data and services to do Geomarketing analyses. The company was founded in 1982. Consequently it was one of the pioneers in the German Geomarketing field. Already 1986 the first version of the Easy Map District Manager was sold. Thenceforth the software was emphasized by a lot of functions. Today Lutum + Tappert provide different attendances from the field of Geomarketing. The main field is the sale and application of their software Easy Map District Manager. With the help of this data of costumers, potentials of markets and data of the population can be visualized per example. Thus a costumized analyses is possible to realize marketing strategies, control the market and check some statistics. Additionally to the software Lutum + Tappert offer data which may be necessary to acquire new costumers or do more detailed analyses. Furthermore the offer services to help companies during their researches.

The most important tool for this thesis is the Easy Map District Manager. This one is provided in two different versions, which included different options and analyses functionality. The Easy Map Standard Desktop Edition is useful to create maps and do some geographically analyses. However the Easy Map District Manager Desktop Edition may necessary to do Area segmentation and planning of locations. Both editions are available as a demo version, consequently a deeper inside into the functionality of the software is possible. After the installing of the District Manager some sample data can be used to do some analyses. Therefore zip code areas of Hamburg and the neighbourhood of it are taken. The tool offers three possibilities to do area segmentation. The first one generates new areas if no distriction was done before. Within the test scenario a given number of areas are created by allocating the zip code areas to districts. The number of zip codes areas within every district should be as balanced as possible compared to the others. After the allocation is done locations are set into the middle of the area to show the position of possible new company sites. These approach is similar to the Greenfield analyses which will be explained later. The second functionality



uses an distribution of areas that already exists. The containing zip code areas in every superior area will be rearranged if it is necessary to get a balanced area segmentation concerning the number of zip code areas within every district. Existing locations within that dataset are not heeded during the rearrangement. The third option uses just existing locations. With the help of these sites the zip code areas are allocated in this way that every locations is placed within one district. Additionally every district should have the same number of zip code areas again. These two provided functions are both from the field of the optimization of area segmentation. In reality often additional data are used during the rearrangement per example the number of households or the purchasing power. These information are used during the area segmentation and are often the parameter which should be balanced. Such a use case was not possible to test because such data are not available for the chosen area. Nevertheless an analyses concerning the number of postcode areas can be done so that an inspection of the results can be done. It is recognizable that the resulting areas are mostly coherent but there is no need to get an absolutely coherent district. Especially at borders to other regions some postcode areas are located inside of the neighboured region. Consequently it is not connected directly to the superior district which it belongs to. Thus the conclusion can be achieved that no checking of coherence will be done. By that reason the islands of the north sea do not lead to some problems during the allocation.

**Figure 7**  
*Creating twenty areas with Easy Map District Manager concerning a balanced number of zip codes areas for each district. The clipping shows the creation of incoherent areas.*



After doing the area segmentation the locations are placed into the middle of the appropriate district. In this case no other placement can be chosen. Just if an



optimization of company sites should be done the locations can be placed into the weighted centroid of the district, too. Furthermore it is recognizable that no new locations can be added to the area where already a certain number of sites exist. But this will be necessary if a new company sites need to be opened. This fact will be described later on with the help of the term Whitespot analyses.

Nevertheless the Easy Map Manager offers a lot of possibilities to do some Geomarketing analyses. With the help of video tutorials on the website of the company working with the software is easy. Consequently no time consuming contraction is necessary.

### 2.3 SIM Tool

## 3 Application of area segmentation

### 3.1 Introduction

inklusive Basisbegriffe, wie base area,... centren gegeben oder nicht, hier untersuchung auf gegeben,da später anwendung auf geomarketing

### 3.2 Parameters

mögliche Bedingungen

### 3.3 previous researches

- Overview of solution techniques - split resolution problem: erkenntnis: However we found that the running time is still too high for the solution of large-scale problems with many thousands of basic areas in an interactive environment. We can make the allocation step much faster by assigning every basic area to the nearest center. This means that we drop constraints (2). This AllocMinDist heuristic has a greatly reduced running time. However, as one would expect and as the computational results in Section 7 show, the balance of the territories obtained is not satisfactory. Therefore in the next section we present a new heuristic based on geometric ideas. It has the desirable property of being very fast (comparable to location-allocation with Alloc- MinDist) and producing territories that are balanced comparable to location-allocation with TRANSP and split resolution with AssignMAX. deswegen versucht andere algorithmen zu finden, Quelle: bericht71.pdf

### 3.4 Use cases

#### 3.4.1 Sales Districting

siehe gebietoptimalaufteilen pdf seite 29

As each territory elects a single member to a parliamentary assembly, the main planning criteria is to have approximately the same number of voters in each territory, i.e. territories of similar size, in order to respect the principle of one manone vote. bericht71.pdf

If unequal territories exist and if it is generaUy known by the salesmen that work load or territory potential is disproportionate, this can lead to low morale, poor performance, a high turnover rate, and an inability to assess the productivity of individual territories or districts. By reawith respect to work load or sales potential, a more optimal utilization of each individual salesman can be achieved. hessandstuart.pdf

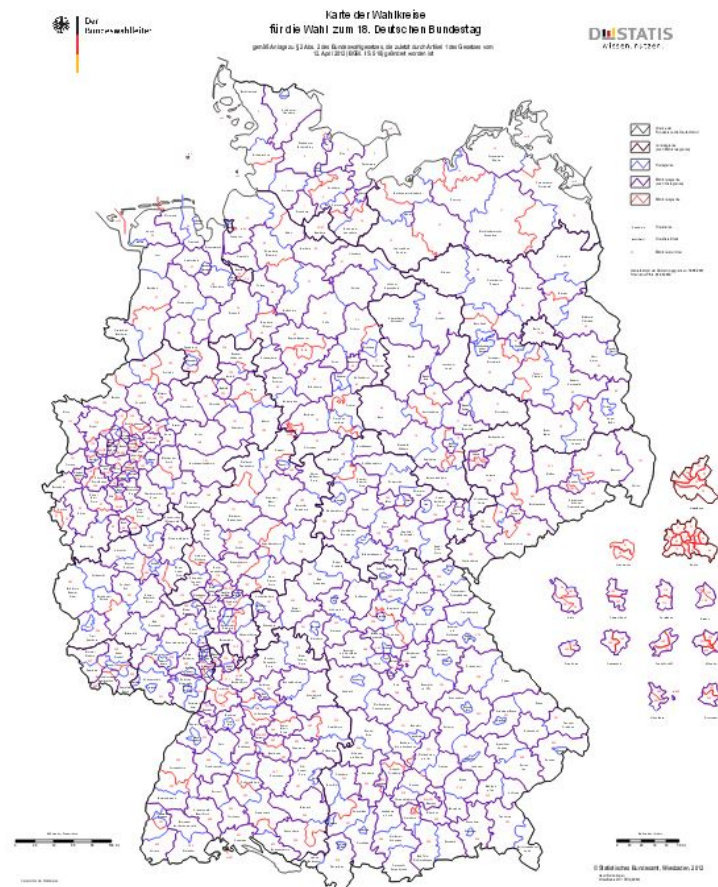
#### 3.4.2 Political Districting

The legislative districting problem the "one man-one vote" problem is to subdivide a state into a specific number of compact and contiguous districts of nearly equal population. Quelle: Hess and Stuart

There are three essential characteristics of districts: the districts should have nearly equal populations to adhere to the one-person, one-vote principle; the districts should be contiguous; and the districts should be geographically compact. Quelle: Mehrotra

Within determined time intervals elections are done within a country to vote for persons who want to be the representatives of a country, federal state etc. Therefore the area has to be divided into sub parts, so called constituencies. Every constituency nominates one candidate who will be elected directly into the parliament. A democratic election is based on the same weight of every voting that is why some restrictions have to be followed during the political districting. In Germany these conditions are set down into the Federal Electoral Law §3 Art. 1. It determines that the creation of constituencies should be done in this way that the number of constituencies within the federal state should be agree with the part of the population [1]. This means that every constituency should hold a similar number of voters compared to other constituencies. The number of inhabitants of Germany is used as stipulation to satisfy that condition. During the districting the boundaries of townships, districts and cities should be preserved as much as possible. Before an election can be carried out the constituencies have to be proofed and adapted if it is necessary because local alteration of the population can be recognized over the time. A commission will do this in front of every voting. The figure below shows the political districting in 2013 for the parliamentary elections.

**Figure 8**  
*Political Districtings in Germany for parliamentary elections in 2013*  
 [1]



3.4.3 Greenfieldanalysis

3.4.4 Whitespotanalysis

## 4 Area segmentation approaches and their implementation

### 4.1 Segmentations considering homogenous distribution

#### 4.1.0.1 Considering just Criteria

#### 4.1.0.2 Considering average criteria value in proportion to number of locations

### 4.2 Segmentations considering distance

#### 4.2.0.3 Considering just Distance

### 4.3 Segmentations considering criteria and distance

#### 4.3.0.4 Criteria and Distance: from inside to outside: SmallestCrit-GetsNearest

#### 4.3.0.5 Criteria and Distance: from inside to outside: SmallestCrit-GetsTrueNearest

#### 4.3.0.6 Criteria and Distance: from outside to inside + inside to outside

#### 4.3.0.7 sum of criteria divided by number locations + Distance

#### 4.3.0.8 Distance + criteria

## 5 Problems of the implemented approaches

### 5.1 Performance

### 5.2 Requirements from the field of Geomarketing

#### 5.2.1 Incoherent areas

#### 5.2.2 Inhomogeneous distribution

beiden subsubsection darunter gehören eigentlich mit zu diesem Gebiet

#### 5.2.3 Rearrangement / Infinite loops during rearrangement

#### 5.2.4 Need of threshold Values

## 6 Comparison of implemented approaches

### 6.1 Performance

### 6.2 Problems

### 6.3 Requirements

### 6.4 Conclusion

## 7 Application of algorithm XYZ to Geomarketing analysis

### 7.1 Optimization of Areas

#### 7.1.1 Conditions and Aims

#### 7.1.2 Approach of algorithm

### 7.2 Greenfieldanalysis

#### 7.2.1 Conditions and Aims

#### 7.2.2 Approach of algorithm

### 7.3 Whitespotanalysis

#### 7.3.1 Conditions and Aims

#### 7.3.2 Approach of algorithm



## 8 Realworld scenario: integretaded Algorithm to the mapChart Manager

## 9 Evaluation

## 10 Discussion and Perspective

### 10.1 Summary

### 10.2 Limitations

### 10.3 Comparison to related work

zu algorithmus von kit vergleichen: vorteil: beste lösung wird ermittelt nachteil: bei viele daten expoentiell großer baum, nur auf gebietsverteilung ohne standorte, um rechenzeit einzuschränken nur limitierte anzahl an richtungen der Linien + nur geraden, somit leidet die Kompaktheit, außerdem verschnitt auf PLZ-Gebiete notwendig

### 10.4 Perspective

## References

- [1] Statistisches Bundesamt Der Bundeswahlleiter.  
URL: [http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW\\_BUND\\_13/](http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/),  
last visit: 30th of October 2015. 17, 18
- [2] J. M. Frühling and W. Steingrube. *Geomarketing: Neue Begriffe = neue Methoden?* Number 39(3-4), S. 184-194. Zeitschrift für Wirtschaftsgeographie, 1995. 6
- [3] GfK Geomarketing GmbH. *Praxiswissen Geomarketing*. Number Bericht 71. Fraunhofer Institut Techno- und Wirtschaftsmathematik ITWM, 2005. 6, 7
- [4] Lutum+Tappert Dv-Beratung GmbH.  
URL: <http://www.geomarketing.de/>,  
last visit: 03th of November 2015.
- [5] J. Kalcsics, S. Nickel, and M. Schröder. *Towards a Unified Territory Design Approach Applications, Algorithms and GIS Integration*. GfK Geomarketing GmbH, 2012. 11
- [6] microm Micromarketing-Systeme und Consult GmbH.  
URL: <http://www.microm-online.de/>,  
last visit: 23th of March 2015, .
- [7] microm Micromarketing-Systeme und Consult GmbH.  
URL: <http://www.microm-online.de/>,  
last visit: 03th of November 2015, .
- [8] microm Micromarketing-Systeme und Consult GmbH.  
URL: <http://www.mapchart.de/>,  
last visit: 03th of November 2015.
- [9] Dennis Müller.  
URL: <http://www.omkt.de/geomarketing/>,  
last visit: 29th of October 2015.
- [10] M. Nitsche. *Micromarketing: Daten-Methoden-Praxis*. Wirtschaftsverlag Ueberreuter, Wien/Frankfurt, 1998.
- [11] KIT Institute for Operations Research Prof. Dr. S.Nickel. Lizard - library of optimization algorithms for distrcting.  
URL: <http://dol.iior.kit.edu>,  
last visit: 03th of November 2015, 2014. 12