

# LocalCompanies: Visual Analytics of spatial aligned regional companies

Alrik Hausdorf\*, Andreas Niekler†, Daniel Wiegrefe\*

\* Image and Signal Processing Group, Institute for Computer Science, Leipzig University, Leipzig, Germany  
E-mail: {hausdorf,daniel}@informatik.uni-leipzig.de

† Natural Language Processing Group, Institute for Computer Science, Leipzig University, Leipzig, Germany  
E-mail: aniekler@informatik.uni-leipzig.de

**Abstract**—Municipal authorities have, among other tasks, a great interest in supporting their local economy. For this purpose, they provide consulting offices that advise companies and mediate cooperation partners.

The city administration of Leipzig created a business register in which companies can provide their competences in free text fields. This business register contains over 1000 entries and it is not straight forward to find and compare companies based on their self-descriptions.

In this paper, we propose a new visualization to analyze the distribution of local companies and exploring the competence profiles of the companies. In order to visualize connections between companies, we perform a semantical analysis. In detail, we use the management staff listing and the core competence descriptions to link the entries. The company location and the connections between the companies are visualized on a map or as a graph. The visualization provides several filtering and interaction mechanisms on demand. From a governance perspective, this leads to insights into company and industry sector networks within a local economic zone.

**Index Terms**—visual analytics, geospatial regional connections, company analysis, heatmap

## I. INTRODUCTION

Promoting the networking of local companies is one of the main tasks of economic development offices of municipal administrations. Therefore, these offices collect and maintain description profiles of local companies to generate a competence profile of each company. Combined with the location of the companies it is possible to explore the local distribution of companies or to find a local partner for a business project. In this paper, we propose a new visualization to analyze the distribution of local companies and exploring the competence profiles of the companies to facilitate the search for local cooperation partners. We use a curated data set provided by the economic development office in Leipzig. The data set contains 1228 companies in total. There is a strong focus on the media business sector with 1043 companies. Therefore, we only concentrate the visualization on this reduced but focused set. To visualize connections between companies in the data set we perform a semantical analysis based on the management staff list and the core competence descriptions. After cleaning and tokenizing, the data is transformed into a document term matrix and a topic modeling is applied. Afterward, we interpret

the inferred topics inductively and assign a label to each topic. In the visualization approach, the user can use these labels and filter the company map at a competence level. Additionally, we map the addresses of the companies to GPS coordinates for the visualization.

We use a tile map (1) as a base layer for the visualization approach and mark every company with a blue dot on the map. The user is able to change the map layout and the city district boundaries are shown for better orientation. Due to the overlap of companies at the same location, we provide a heat map aggregation to visualize the distribution of companies over the city area. The user can search for specific companies by filters, e.g., using a text search or filtering for the topic clusters. To gain more detailed information about a company, an information box is provided next to the map visualization on demand by clicking the data point on the map. This information panel also shows the calculated network information (3) if there are overlaps with other companies. By selecting a node from the network visualization the user can filter for this company in the map visualization.

The visualization was evaluated with an interview of an expert, who works at the economic development department of the city of Leipzig. In particular, the explorative search options were considered to be very useful to reveal previously unknown connections between companies.

## II. RELATED WORK

Common GIS tools e.g. ArcGIS can show spatial data to experts and let them analyze them. Mostly they need already calculated coordinates and do not provide easy calculation systems for that. As shown by Lu et al. [1] GIS tools like WebGIS have many complex ways to display data and have to need to be heavily configured for non-academic purposes.

Different systems are available to show geo-encoded data on a map. For example, the tool GeoTemCo [2] uses an additional temporal dimension to analyze high amounts of geo-encoded data over time. It aggregates the data by their location to reduce overlapping. Without the temporal dimension and the advantages of the aggregation, this tool is not suitable for our purpose.

With the release of web-based maps like Google Maps [3], Apple Maps [4], Bing Maps [5], or Open StreetMap [6] users

are already familiar with geolocated points. Lan et al. [7] are using a different type of glyph to show events, extracted from news and twitter messages, located on a map. This visualization type uses single points of interest located on a map. If the data consists of a high amount of (nearly) overlapping points, one can hardly distinguish between the shown data points.

To overcome this problem overlapping data can be aggregated to form areas. One idea is to calculate representatives of a specific area and place that label on the map. That technique is used by Reckziegel et al. [8]. Their method aims to give a rough overview and some information is lost in the process.

The development of map-providers mostly results in an API that can be used for different geospatial tasks, like the translation of an address to a longitude and latitude coordinate. OpenStreetMap is the only provider that allows free and unlimited access. Furthermore, it is possible to set up their nomination service to translate addresses into GPS coordinates on self-hosted hardware. This service can be used easily to translate high amounts of addresses. For academia, BingMaps provides a special license for usage and delivers more accurate results of the address translation.

The usage of heatmaps which show a distribution of single points on a map is a common task and has been evaluated several times. For example, it was used by Maciejewski et al. on different multiple geo-spatiotemporal datasets [9]. This type of location aggregation is used mostly in order to gain an overview of the data. Therefore several vast-challenge papers used the technique for spatial data ([10], [11]).

### III. METHODS

Due to the number of companies in larger cities, it can be challenging to get an overview of the variety of companies in this city. Furthermore, it can be tedious to find a connection between these companies since it is mostly a manual task for interested users. Therefore, we present a tool to improve the exploration of local companies by their main competence. We show how companies are connected by the same competences and we help to analyze the connections between companies by the same persons in management.

Our processing chain reads the given data set (see subsection III-A) and applies different preprocessing steps (see subsection III-B) to extract any necessary data used in the GUI described in subsection III-C. To handle the data and allow editing of the single records we use a basic RESTful web-service written in JavaScript with the nodeJS framework.

#### A. Data set

The data set of companies is manually curated by the regional business development department of the city of Leipzig. Each change of a data record is entered manually by an employee of the city and the companies can not edit their information. The data set contains 1043 companies related to the media and creative business sector on which we are focused in this paper. These companies are located at 870 unique locations and only 5 companies do not have

a valid address. Complete locations contain the street name and number, the postal code and the city. The data set is not restricted to the city of Leipzig and contains some entries from other cities nearby. The companies are manually assigned to finer business sector categories. Their distribution is shown in Figure 2.

#### B. Preprocessing

After importing the data set we extract the unique addresses to translate them into GPS coordinates. This task was performed using the academic licensed version of the "Find a Location by Address" of Bing-maps Rest Services [12]. All requests are saved in a cache which prevents the tool of requesting the same address multiple times. Afterwards, the coordinates are added to the corresponding data records of the companies.

To visualize connections between companies in the data set we perform a semantical analysis based on the management staff list and the core competence descriptions. We annotate our data set, and finally the visual interface, with information about competence and management overlaps. From a governance perspective, this leads to insights into company and industry sector networks within a local economic zone. We extract the manager information for each company by separating single names from the data field. In addition to the manager information, we apply a linguistic preprocessing to the company competence description fields including the processing sequence of

- 1) removing numbers, punctuations and symbols,
- 2) transformation to lowercase characters,
- 3) stopword removal and
- 4) stemming (The linguistic preprocessing and topic modeling was performed using R and the libraries `quanteda`, `topicmodels` and `dplyr`).

The text source is tokenized and transformed into a document term matrix (DTM). We further filter the matrix to reduce the noise for the topic modeling process. Only terms are kept which appear in more than one document. After the transformation of the texts into a DTM, we apply the classical LDA topic model algorithm [13].

In total, we experimented with topic numbers from 4 - 20 topics. With the 6 topics, we reproduce the intended separation of profession fields best as no redundant topics were left and we have no two professions encoded in one single topic. As a result, each competence entry, each company respectively, is modeled into a multinomial distribution over the 6 topics. The predominant topic for each description can be identified by ordering the multinomial distribution and identify the topic with the highest probability share. In this way, the most important topic for each company competence description can be identified. Finally, we build a list of all topics (The topic names are given by the context of the most probable words in the topics) and assign the company descriptions to the topics. Additionally, we interpret the inferred topics inductively and assigned a label to each topic. The resulted topics can be found in Table I. In the visualization approach, the user can use

these labels and filter the company based on the extracted competency skills. This data set is merged back into the data basis and can be used for the visual interface. In effect, we now have a single topic ID augmenting a company in the data set. The topics can be displayed by their topic name and companies are identified by the topic id. The manager names are matched by a direct lexical comparison, where only 100% matches are taken and linked.

TABLE I  
LIST OF THE CALCULATED CLUSTERS THAT CONNECT COMPANIES BY THEIR CORE COMPETENCE DESCRIPTIONS

Clustername (german)	Clustername	#
Verlage	publishers	183
Druck, Herstellung	print, production	186
Software, Service	software, service	175
Marketing, Kommunikation	marketing, communication	226
Individuelle Kundenlösung	individual customer solutions	117
Film, Produktion	film, production	154

### C. Visual Interface

The user interface was created as one-page JavaScript web-application using the framework vueJS. It consists of two main parts, the main view with a map to explore the data and a data management view to update the data.

1) *Main View*: The main view of the application uses the enriched data to draw an overview map, apply filtering and perform search tasks. Furthermore, it gives the user detailed information about specific companies. These view contains the following areas:

**a) Map area** This area contains a layer-based map and different tile-based layers can be used as an underlying map to the displayed data. The next layer displays blue-transparent areas in the different granularity of the region separations. Another layer is the heatmap-layer that can be used to gain an overview of the distribution of the selected companies. On top of these layers, the point-layer is rendered. Each point represents a company located by its address, where each of the points is clickable and if the click is unique to one company this company is selected for the detail-view as seen in Figure 1. If the selected point is connected to more than one company or more companies are located at the same address the GUI lists all available companies at that point in the lower-left edge of the map and waits for the selection of one of them. The map areas start with zoom and positioning that shows the area of Leipzig as an overview and the user can explore the data set with standard interactions like zooming and hovering the map.

**a.1) Map filter** The filter area is hidden by default and can be activated on demand by the user. The activated version is shown in Figure 2. It contains two areas to filter the shown companies: The top area represents the manually curated categories contained in the data set. It provides a bar chart to visualize their portion in the data set. The length of the bar represents the number of companies represented in the category in relation to the maximum of companies in all

shown categories. Each category can be toggled to be shown or hidden from the map area. The bottom area shows the filter options. Each entry represents a filter that can be toggled by the button on the right. Currently, filters for the management staff connections and the extracted competencies are implemented.

**b) Search bar** This area allows filtering companies by a string search on the company name. With the help of a typeahead modal, the search can be auto-completed to one company.

**c) Cluster selection area** This area helps to filter by the computed competence clusters (see subsection III-B). Only one competence cluster can be selected simultaneously.

**d) layer selection area** With this option, the user can select different base type layers for the map visualization or add an additional layer that is projected on the map. The possible Layers are listed in Table II.

TABLE II  
LIST OF ALL LAYERS THAT ARE AVAILABLE IN THE MAP VISUALIZATION. EACH LAYER IS IDENTIFIED BY ITS NAME AND IS GROUPED BY TYPE. THE ORDER OF THE LAYERS DETERMINES THE DRAWING ORDER IN THE MAP VIEW.

Name	Type	enabled by default
Stamen Watercolor	base	no
Stamen Terrain only	base	no
Open Street Map	base	no
OSM no labels	base	yes
Regionen	additional	no
Ortsteils	additional	no
Stadtbezirke	additional	yes
Company-Heatmap	additional	yes
Company-Dots	additional	yes

**e) Company detail area** This region shows details on demand for a company that was selected by the user on the map. The company information is separated into a general part in the top, the management staff list below and a part that describes the company by its extracted competences.

**f) Manager connection area** Below the company detail area the connections to other companies are displayed using a force-directed-layout graph layout. The red highlighted dot shows the currently selected company. Each edge of the graph indicates a connection introduced by a person that is present in both companies. With the top right located toggle-button, a filter for the map can be applied to filter only for the involved companies.

**g) Description connection area** Equivalent to the manager connections this area represents the connections that are based on the extracted competences. If companies are part of the same competence cluster, they are connected by edges. Since this connection occurs much more often than management staff list connections and the resulting graph is significantly larger, it is shown in a separate popup on-demand as shown in Figure 3.

2) *Data Management*: To enable the user to change the data set we implemented a data management view. It consists of two parts that help to find and edit single companies.

**Table View** The user starts with a table showing all avail-

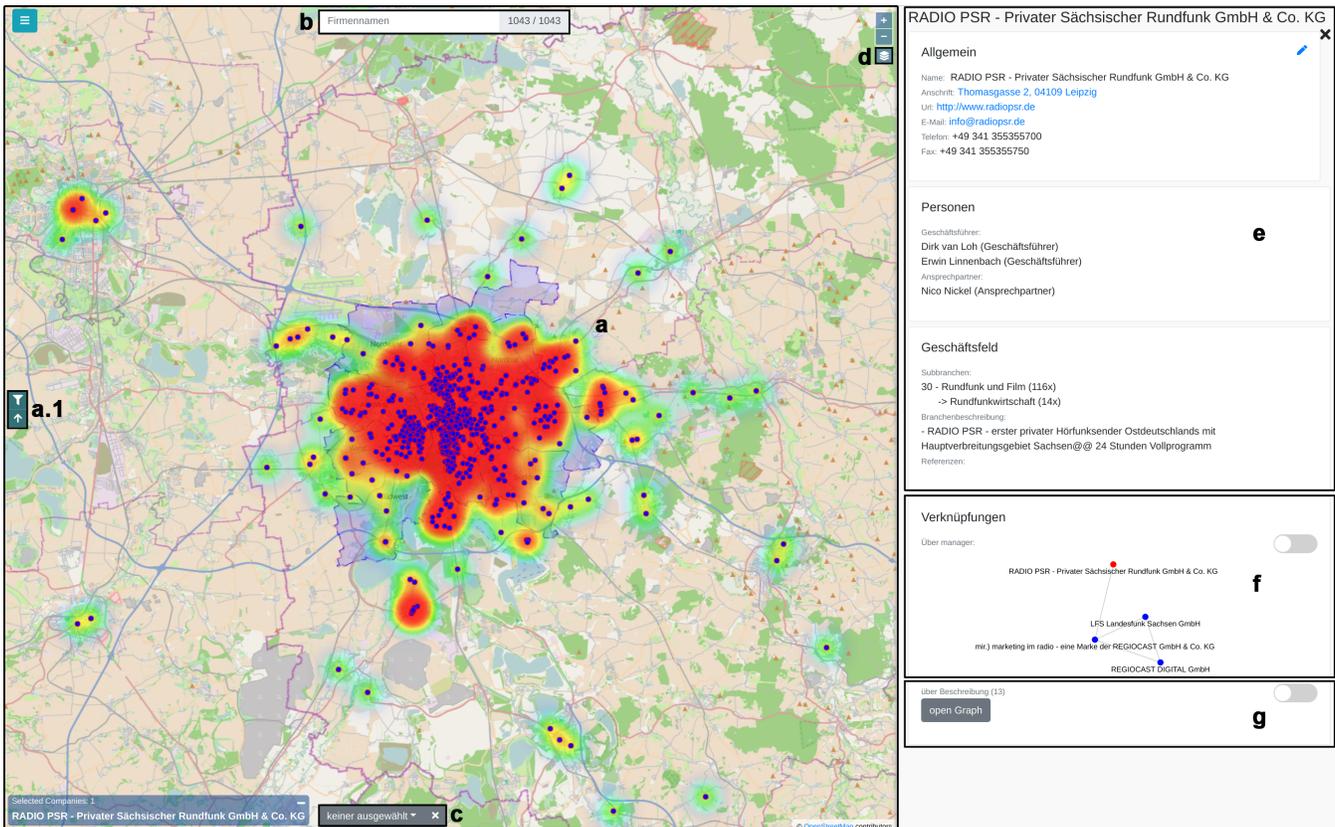


Fig. 1. The main user interface of the tool consists of the map area (a) with filters (a.1), the search bar (b), the cluster selection area(c), the layer selection area (d), the company detail area (e) with the manager connection area (f) and the description connection area (g).

able companies in the data set. The table contains different information per entity. The first column contains the id of the company and is used for the initial ordering of the list. Next, the second column contains the name of the company and the user can select to be the ordering instance for the list. Additionally, a full-text search field is provided to find a specific company by its name. In addition, the third column contains the address. There is an aggregated field containing the street name followed by the street number, the postcode, and the city name. Furthermore, the next column indicates the origin of the coordinates. The differentiation if the coordinates are computed or manually assigned and/or of what origin they are is depicted by a glyph. It is possible to filter the elements by the type of origin. The shown example in Figure 4 indicates an automatic assignment of the coordinates by the Bing API. The last column contains a button which opens the detailed view of the corresponding company. The amount of currently shown elements, as well as the number of total available companies, are shown in the top left area of the view.

**Detail View** This view is used to show all available attributes of a company-entry and enables the user to edit them. These attributes are grouped into the following tabs: "General", "Contact", "Management" and "Geo". Each tab contains the respective attributes. The general tab (see Figure 5) provides the attributes "Branchenbeschreibung" and

"ProdDLBeschreibung" together with a preview. This preview has been proven to be useful since the data set contains the symbols "##" as line breaks. All other fields can be edited with a text field.

The contact tab (see Figure 6) shows the contact information of a company.

The management tab (see Figure 7) contains attributes for the managers and the contact person.

In the last tab (see Figure 8) the geo information are grouped. The tab content is separated by two areas: The top area is similar to the previously used tabs and the user can edit the textual geographical information like the street name. In the bottom area, the functional information about the location is editable. First, the user is able to enable or disable the usage of the geographical location of the company in the visualization. The geo confidence score is a value between 0 and 1 that indicates how accurate the longitude and latitude value was predicted. Different nomination APIs provide the translated GPS coordinates with this value. If a user edits the location manually, the confidence value should be set to 1. Furthermore, the user can modify the origin as well as the value of the longitude and latitude. On the right side, a map with the shown coordinates is provided. For example, the user can use other search engines to look up the GPS coordinates for a specific address and update them in the data set.



Fig. 2. The filter window contains two areas: The top one can be used to filter the displayed companies in the map-area by the manually assigned categories. Each category can be selected to show or hide the corresponding companies. Below every element, we include a bar graph which shows the number of companies in relation to the maximum of companies in all categories. In the bottom are toggle buttons which can filter companies that have connections by managers or the description.

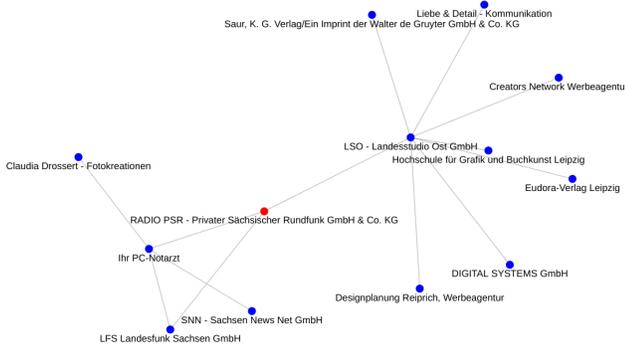


Fig. 3. The description network that is created by the descriptions of the company scope. The red highlighted dot represents the currently selected company. Each edge shows a high similarity between the description of the company scope.

After all changes, the user has to save the changed data set.

#### IV. EVALUATION

The tool was shown to the regional business development department of the city of Leipzig. They found it useful to see those hidden connections. The interface was described as useful and easy to understand. They bring up the wish to make that work publicly available on their platform and make it usable for everyone.

#### V. CONCLUSION

In this work, we demonstrate a tool that helps to understand the regional distribution of companies from different areas of interest. In combination with the manager connections and the descriptions, we give an insight into the data which would be hidden without this data linking technique.

We assume that different areas of industry prefer different areas of the city. In a following test, we want to work with a bigger amount of companies and will test if the visualization and analysis task highlights industry clusters and scales well with larger data sets.

One idea is to use open databases about companies and their management to populate our data set with more connections and companies. Furthermore, it might be useful to extract the addresses of the companies from their business websites to improve the accuracy of the GPS coordinates.

The heatmap approach is able to handle a large number of data records. However, the GUI would fail to show larger data sets as single data points on the map due to the huge load for the web browser. Therefore, it is necessary to develop different aggregation types to handle this known shortcoming of our approach.

#### VI. ACKNOWLEDGMENTS

This research was supported by the city of Leipzig, the office of economic development. This research was funded in parts by the Development Bank of Saxony (SAB) under project number 100335729.

#	Name	address	Actions
244	mir.) marketing im radio - eine Marke der REGIOCAST GmbH & Co. KG	Thomasgasse 2, 04109 Leipzig	→ open
279	Radio Leipzig BCS Broadcast Sachsen GmbH & Co. KG (Leipzig 91.3)	Friedrich-List-Platz 1, 04103 Leipzig	→ open
366	mephisto 97.6 - Das Uni-Radio	Universitätsstraße 3, 04109 Leipzig	→ open
787	RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG	Thomasgasse 2, 04109 Leipzig	→ open

Fig. 4. Table view of the data management part. Filtered by the term "radio" in the company name and ordered ascending by the id.

## RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG

General **Contact** Management Geo Save

ID: 787

Freigabe: -1

Name: RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG

Branchenbeschreibung: - RADIO PSR - erster privater Hörfunksender Ostdeutschlands mit Hauptverbreitungsgebiet Sachsen@@@ 24 Stunden Vollprogramm

Branchenbeschreibung (Vorschau): - RADIO PSR - erster privater Hörfunksender Ostdeutschlands mit Hauptverbreitungsgebiet Sachsen@@@ 24 Stunden Vollprogramm

ProdDLBeschreibung: - privater Hörfunksender mit 24 Stunden Vollprogramm

ProdDLBeschreibung (Vorschau): - privater Hörfunksender mit 24 Stunden Vollprogramm

intern\_AufnahmeDatum: 02.09.2009

intern\_LetztesDatum: 12.11.2015

wz2008: 601

MKWintern: 30

MKWinternName: Rundfunk und Film

MKWinternSub: Rundfunkwirtschaft

Fig. 5. The general tab of the detail view of the data management part.

## REFERENCES

- [1] H. Lu, W. Nihong, W. Chang, and C. Yujia, "The research on the webgis application based on the j2ee framework and arcgis server," in *2010 International Conference on Intelligent Computation Technology and Automation*, vol. 3, May 2010, pp. 942–945.
- [2] S. Jänicke, C. Heine, and G. Scheuermann, "Geotemco: Comparative visualization of geospatial-temporal data with clutter removal based on dynamic delaunay triangulations," in *Computer Vision, Imaging and Computer Graphics. Theory and Application*, G. Csurka, M. Kraus, R. S. Laramée, P. Richard, and J. Braz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 160–175.
- [3] "Google maps," Website, <https://www.google.de/maps>; accessed on 01-11-2019.
- [4] "Apple maps on the web," Website, <https://developer.apple.com/maps/web/>; accessed on 01-11-2019.
- [5] "Bing maps - directions, trip planning, traffic cameras & more," Website, <https://www.bing.com/maps/>; accessed on 01-11-2019.
- [6] "Openstreetmap deutschland: Karte," Website, <https://www.openstreetmap.de/karte.html>; accessed on 01-11-2019.
- [7] R. Lan, M. D. Adelfio, and H. Samet, "Spatio-temporal disease tracking using news articles," in *Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health*, ser. HealthGIS '14. New York, NY, USA: ACM, 2014, pp. 31–38. [Online]. Available: <http://doi.acm.org/10.1145/2676629.2676637>
- [8] M. Reckziegel, M. F. Cheema, G. Scheuermann, and S. Jänicke, "Pre-dominance tag maps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 6, pp. 1893–1904, June 2018.
- [9] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert, "A visual analytics approach to understanding spatiotemporal hotspots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 205–220, March 2010.
- [10] C. Scruggs, C. Henkel, and C. D. Stolper, "Audio explorer - vast challenge 2018 mc1 (awarded excellent comprehensive submission)," in *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Oct 2018, pp. 92–93.
- [11] A. Rukavina and S. Banchero, "Vast challenge 2018: Mini-challenge 1 award: Applied visual data science," in *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Oct 2018, pp. 90–91.
- [12] "Bing maps rest services," Website, <https://docs.microsoft.com/en-us/bingmaps/rest-services/>; accessed on 01-11-2019.
- [13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

**RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG** Save

General Contact **Management** Geo

E-Mail1	info@radiopsr.de	E-Mail2	
Url1	http://www.radiopsr.de	Url2	
Tel1	+49 341 355355700	Tel2	
Fax1	+49 341 355355750	Fax2	

Fig. 6. The contact tab of the detail view of the data management part.

**RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG** Save

General Contact **Management** Geo

GFFunktion	Geschäftsführer	GFVorname	Dirk	GFNachname	van Loh
GF2Funktion	Geschäftsführer	GF2Vorname	Erwin	GF2Nachname	Linnenbach
AspFunktion	Ansprechpartner	AspVorname	Nico	AspNachname	Nickel

Fig. 7. The management tab of the detail view of the data management part.

**RADIO PSR - Privater Sächsischer Rundfunk GmbH & Co. KG** Save

General Contact **Management** **Geo**

Strasse	Thomasgasse
Hausnummer	2
HausnummerZusatz	
Plz	04109
Ort	Leipzig
geoDisable	<input type="checkbox"/>

geoConfidence

geoOrigin Auto: Bing-Maps

geoLat 51,33998

geoLong 12,37401

Ort2 Thomasgasse 2, 04109 Leipzig @ BING-Maps

BingGeo



Fig. 8. The geo information tab of the detail view of the data management part.