Solid Freeform Fabrication 2023: Proceedings of the 34th Annual International Solid Freeform Fabrication Symposium – An Additive Manufacturing Conference
Reviewed Paper

Usage of Unconventional Data Sources for Market Intelligence (MI) in the Field of Additive Manufacturing (AM) - Expert Networks, Technology Territories and Trends

S. Ziegler[1,2], J. H. Schleifenbaum[1,2]

[1] Chair for Digital Additive Production DAP, RWTH Aachen University, Aachen, Germany
[2] Aachen Center for Additive Manufacturing ACAM GmbH, Aachen, Germany

Abstract

The geographic expansion of the markets for AM increasingly confronts companies with greater competitiveness due to the globalization. In addition, market participants are facing rapid changes in the business environment - due to new information and communication technologies. Companies only have a chance to hold their market position if they quickly adopt market changes. Therefore, the decision-making process needs to be accelerated by on-demand information provision. MI offers one possibility to meet these requirements, but typically based on external unstructured data for market and competitive evaluation, which makes it cost and time consuming. A specific investigation of such data sources related to MI for systematic use within the AM markets is being carried out. For this purpose, different data sources (e.g. LinkedIn) will be identified, analysed with focus on information synthesis using text mining and their suitability for the evaluation of expert networks, technology territories and trends be presented.

Keywords: Additive Manufacturing, Trend Analysis, Market Intelligence, Text Mining

Introduction

In the context of globalization, companies must deal with increasing competitive dynamics as a result of the geographic expansion of markets due to growing competition in old and new markets. In addition, market participants are subject to rapid change in the business environment - especially due to the fast development of new information and communication technologies. As a result of the complexity of the increased competitive dynamics, management is required to adjust more and more frequently in order to secure the company's earnings and existence in the long term. Companies only have a chance of holding their position in the medium and long term if they succeed in quickly adapting to the fast-changing corporate environment. Instead of focusing on existing competencies and services, which have been provided efficiently in the past, constant adaptation comes into focus.

New developments, which must be absorbed and processed ever more quickly, put the flexibility of companies and thus also the responsiveness of their employees to the test. Technological innovations and fast-growing data volumes have led to a dramatic increase in requirements for the production factor information and its efficient processing and distribution.

The Market Intelligence (MI) approach provides a way of fulfilling the growing requirements in terms of the need-based provision of information. As an analytical information system, MI typically relies on external, unstructured data to assess the market and the competitive situation, among other factors. In practice, investigations with external, unstructured data sources are usually carried out manually and the information often must be gathered from several sources, as a central, unified database, such as a data warehouse, does not exist here. Consequently, such executions pose some challenges and lead to extreme efforts.

According to the previously described challenge and the associated deficits of the practice, the aim of the present work is to investigate such data sources in the context of MI for systematic use in the industrial environment. For this purpose, different data sources are identified, investigated and evaluated with respect to their use case and the possibilities of automated information retrieval, processing and analysis by means of text mining.

Basics

Market intelligence is defined as the information relevant to a company's market, such as trends, competitors and customers, which is collected and analyzed for accurate and secure decision-making when defining strategy in areas such as market penetration strategy and market development. Data required for this purpose can be collected from primary sources (including surveys, observations) or secondary sources (including external databases, studies and technical literature). For decision making, data quality and quantity should always be considered in addition to interpretation of results. Currently, a lot of tools (e.g., Powder BI) and services (e.g., Echobot, AlphaSense) are available that can provide tailored information from the market based on the decision maker's question. For example, Questback, offers customer and employee survey tools and services, as well as market research, to provide targeted insights for process and product improvements. In addition, further information can be generated on market alignment and development of new markets. Google Analytics, although it is not a technical research tool, provides statistical information on the behavior of customers in terms of their response to a company website. The line between market and business intelligence becomes quite thin in this case, as the insights come from the company's own website, but the triggers for these insights are from the external environment. In this way, a company can decide which product is viewed most often but never purchased. [7]
Market data, e.g., from IFH Cologne, provides access to brand-specific key figures and financial ratios for (sub)markets or industries. This can be used to derive developments and trends and reduce investment risks. Social networks such as Facebook are among the most important data sources for MI today. In addition, influencers, power users and similar individuals can be used to target and influence markets, customers and products. Companies such as kred, khoros or PeerIndex [3] offer the necessary software solution for identifying and managing social media. Based on the origin of the data, a distinction is made between classic market intelligence based on external data, MI that provides insights using information from social media, and business intelligence based on a company's proprietary information. [1] The challenge of the classic method is to track down and analyze the right data sources. This includes information on competitors and their market position. In this way, companies analyze their strengths and weaknesses and identify potential for new developments and market gaps. Some companies specialize in the collection and analysis of such data sources across industries. [1] Competitive intelligence represents a special form of MI that specializes in collecting and analyzing information about competition in the form of news, feedback from sales or suppliers. This can be used to generate in-depth information on market share and positioning, contracts signed, and other information relevant to future market positioning and pricing. [1] Nowadays, social media is increasingly used as a data source for MI. Twitter, Facebook and other social media platforms can be used as data sources for MI e.g. to identify consumer behavior. This can provide companies with better insights to introduce new products or specific improvement proposals from relevant target groups. [1, 2] To verify the informative value of MI data sources, they can be specifically correlated with internal company data sources (business intelligence). For example, it can be determined which new products or customers promise the most profit based on the social media reaction to a product announcement. This information can be cross-referenced with historical internal company data (including sales volumes) on similar products to place the social media reaction in an overall context and evaluate it. [1, 4]

Use case: Identification of experts and technology regions

The increasing globalization and internationalization of markets in recent years offer companies both advantages and disadvantages. Open and transparent markets enable companies to become active in new markets by themselves, but with the disadvantage that this opportunity is also available to companies that were previously outside the market. Another complicating factor is that these new markets require greater segmentation of customer groups, which must be served in a more individualized manner in order to generate competitive benefits. The dynamics regarding changes in product characteristics, technologies and customer requirements are manifested in constantly decreasing cycle times (innovation cycles, product life cycles) with a simultaneous shortening of the half-life of knowledge.

To meet these challenges, a variety of information sources must be consulted early on in strategic decision-making to ensure competitiveness. Strategic issues for a company in the area of market and technology development include:

a. At what level of technology maturity is the industry-specific promising manufacturing technology?
b. At which locations are most of the technology development projects and associated experts located?
c. How are relevant new or existing markets developing?

To answer these questions, various data sources are consulted, automatically analyzed, interpreted and the potential added value is shown based on following use case.

Data Acquisition

To answer the questions, patents, company trade records, LinkedIn profiles and scientific publications are crawled via Amadeus (from Bureau van Dijk), LinkedIn and Scopus (from Elsevier) (see Figure 2). For automated data retrieval, the application programming interface (API) of the respective platform is used and integrated into RStudio. For data reduction, the search is limited to the keyword AM and to the regions Netherlands, Belgium and Germany in the respective database.
In response to the search string, consisting of keyword and regions (see Figure 2), the sub-information (see Figure 1) is returned to RStudio as a data frame.

Information	Sub-Information
Patents	Title of publication, Granted (Yes/No), Date of publication, Current owner(s), Inventor(s) Country code
Companies	Name of the company, location, country code, number of employees, year of foundation
Profiles	LinkedIn-Profile, Name Company, Job Title, Job Description, Job Location, Job Date Range, Company 2, Job Title 2, Job Description 2, Job Location 2, Job Date Range 2, All Skills
Publications	Authors, Title, Year, Source title, Link, Affiliations, Abstract, Author, Keywords

Figure 1: Overview of the contained sub-information

Subsequently, the sub-information contained in the data frame is cleaned of numerical values, punctuation marks, umlauts and stop words that are not to be taken into account during full text indexing (see Removal Function, Figure 2). Likewise, upper case is converted to lower case, manually defined stop words are removed and synonyms for the keyword, e.g. "3d printing", are homogenized. To evaluate the quality of the data frame, it is checked whether the keyword occurs per entry.

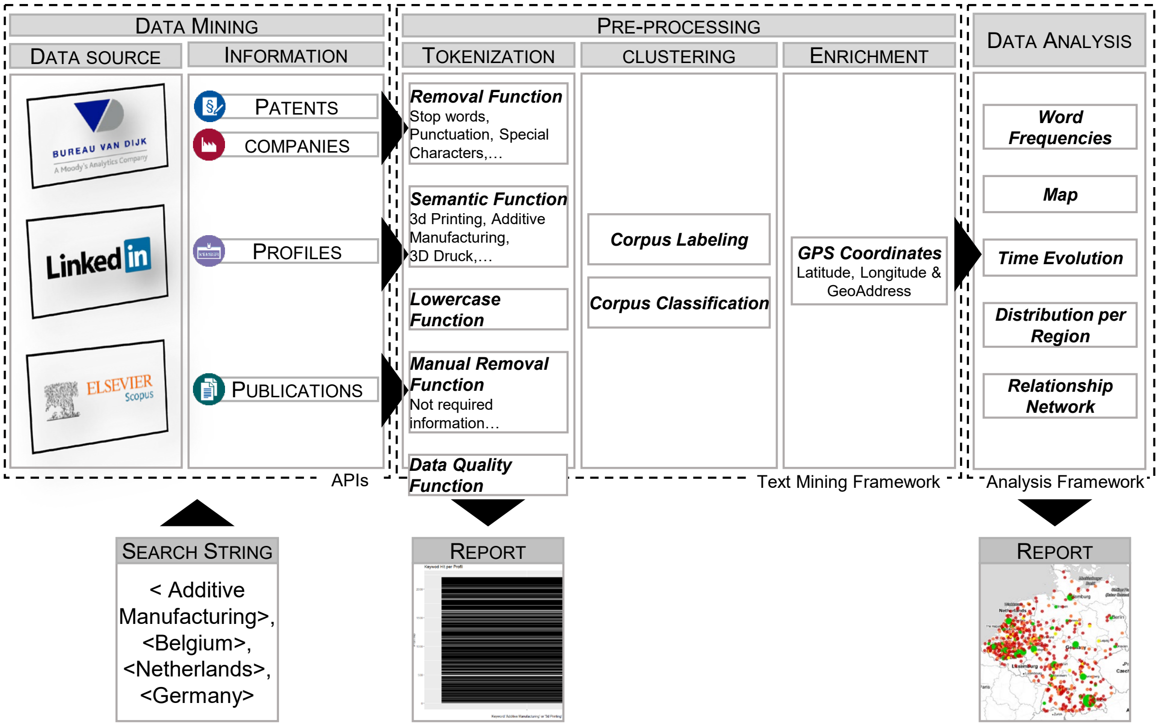


Figure 2: Procedure for data collection and cleaning

For further processing of the data frame, the next step is the classification and labeling of the text corpus contained in the data frame. The enrichment of the sub-information with geo-coordinates is done via the API of Google. Finally, the data is analyzed with regard to the mentioned questions.

Data Analysis

First, the analyzation of the development of publications, patent applications, as well as company start-ups for the period from 2010 to 2019 in the regions of Germany, Belgium and the Netherlands (see Figure 3).

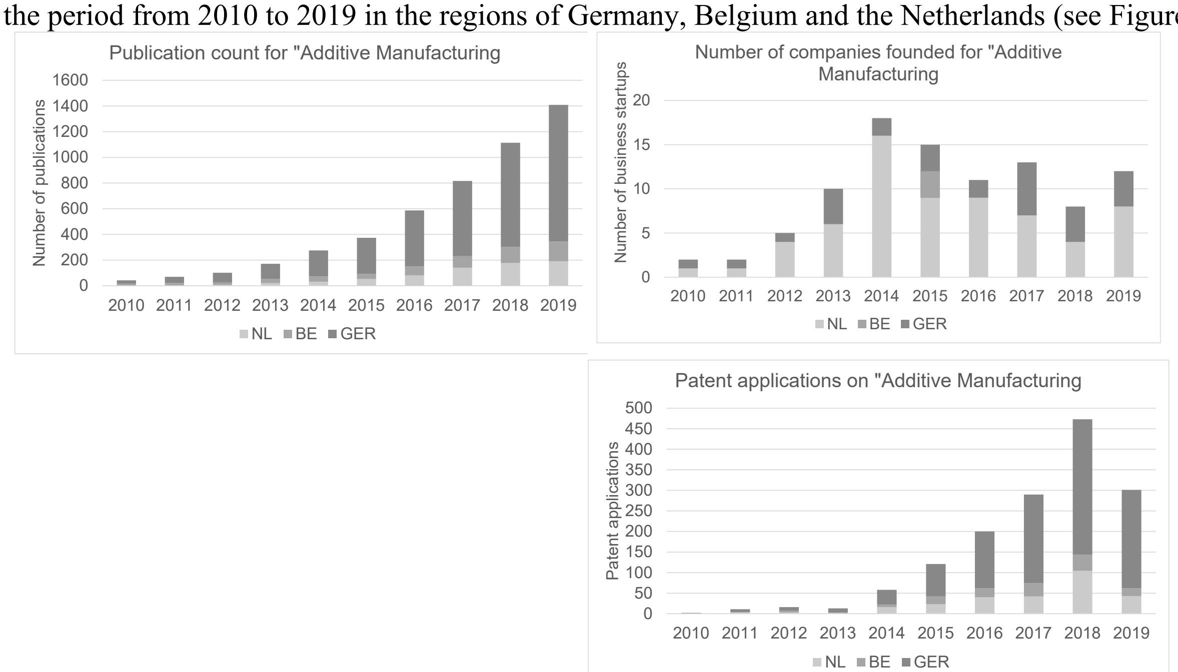
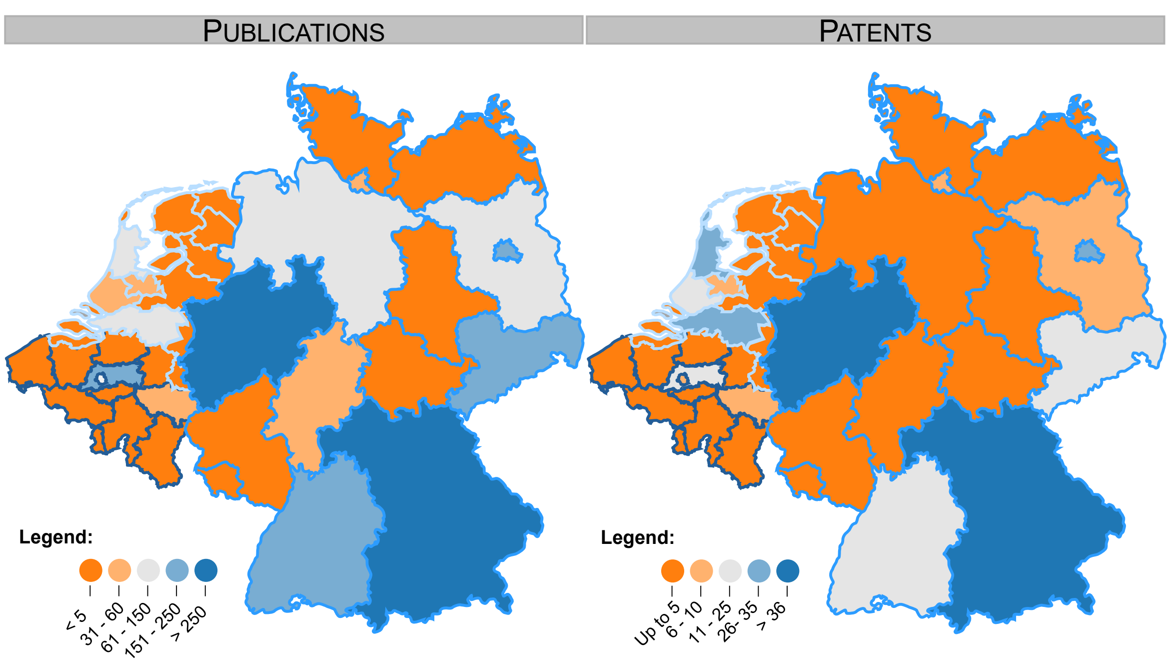


Figure 3: Development of publications, patent applications and start-ups in Germany, the Netherlands, Belgium for the period from 2010 to 2019

The development of publications since 2010 shows exponential growth in all three countries, with most publications in Germany. Thus, research in AM with broad impact has been taking place at German research institutions for several years.
The number of start-ups is highest in the Netherlands compared to the other countries, reaching a maximum in 2014. One possible explanation lies in the tax system in the Netherlands. All income earned by a Dutch parent company from a foreign subsidiary is exempt from Dutch tax. This means that more companies can be established in the Netherlands, which then operate primarily through subsidiaries abroad. Similarly, no distinction is made between new branches with a local legal form and actual independent company formations. Both of the aforementioned explanations also lead to a distortion of the analysis shown.
In the case of patent applications, exponential growth can be observed up to 2019, with Germany having the most applications in quantitative terms. The decrease in patent applications can be attributed to the weak economy of EU member states and the emerging COVID-19 pandemic.
Based on the trend of publications and patent applications from 2016 to date, it can be concluded that there is a greater technology maturity. Likewise, increased industrial use of AM manufacturing technologies is also emerging. This is also corroborated by the industry revenue, which has more than doubled compared to 2015 (as of 2019: approximately $10,100 million) [8] and by the investment of $1.45 billion (in 2019). Supported by this information, a strategy for technology chains in manufacturing can be elaborated, for example, "technological leadership" or "technological presence" [5]. By correlating with industry sales and investment numbers, further statements can be made about market development. Industry sales can be approximated by the sales of companies that are exclusively active in AM. This means that the market development can be described in terms of tendency, but no absolute figures can be calculated. The sales of these companies can also be obtained via Amadeus, but are not considered further in this paper. On the other hand, the investment cannot be recorded via the available data sources.
In the next step, the distribution of publications, patent applications and start-ups among the (federal) states (see Figure 4) and geocoordinates (see Figure 5) of Germany, the Netherlands andBelgium is analyzed in more detail for 2018 as an example. The focus here is increasingly on the evaluation of technology regions and the corresponding experts.



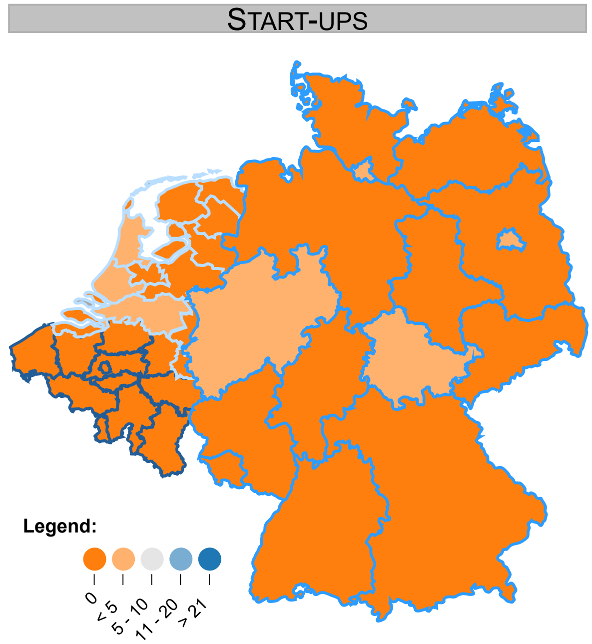


Figure 4: Distribution of publications, patent applications and start-ups among the (federal) states of Germany, the Netherlands, Belgium in 2018.

When comparing the individual countries, it can be seen that Vlaams Brabant (Belgium), North and South Holland (Netherlands), North Brabant (Netherlands), North Rhine-Westphalia (NRW), Baden-Württemberg, Bavaria, Brandenburg, Hamburg, Berlin and Saxony (all Germany) have an above-average distribution in publications, patent applications and start-ups in 2018.

Three approaches now exist for directly comparing the states with each other:

The figures determined for publications, patent applications and start-ups can be normalized by relativizing the key figures using the number of inhabitants per square kilometer and the education index [6] or impact factor. The impact factor is intended to serve as a dimensionless indicator of the scientific quality of the country or region in question, while the education index is used to measure economic development and quality of life. This information can be obtained from various platforms (including Statista, Scopus, UNDP).

The second option is to choose a finer breakdown into (urban) counties instead of the used breakdown into the individual (federal) states. In this way, the individual regions can be better compared and evaluated directly with each other. With both variants, the identification of experts is not directly possible.

The third variant allows a direct comparison of different locations with each other as well as the identification of experts. Due to the granularity of the information, it makes sense to use this approach with one of the two previous variants for rough and detailed planning. The following figure (see Figure 5) shows the distribution based on the geocoordinates for Belgium, the Netherlands and Germany.

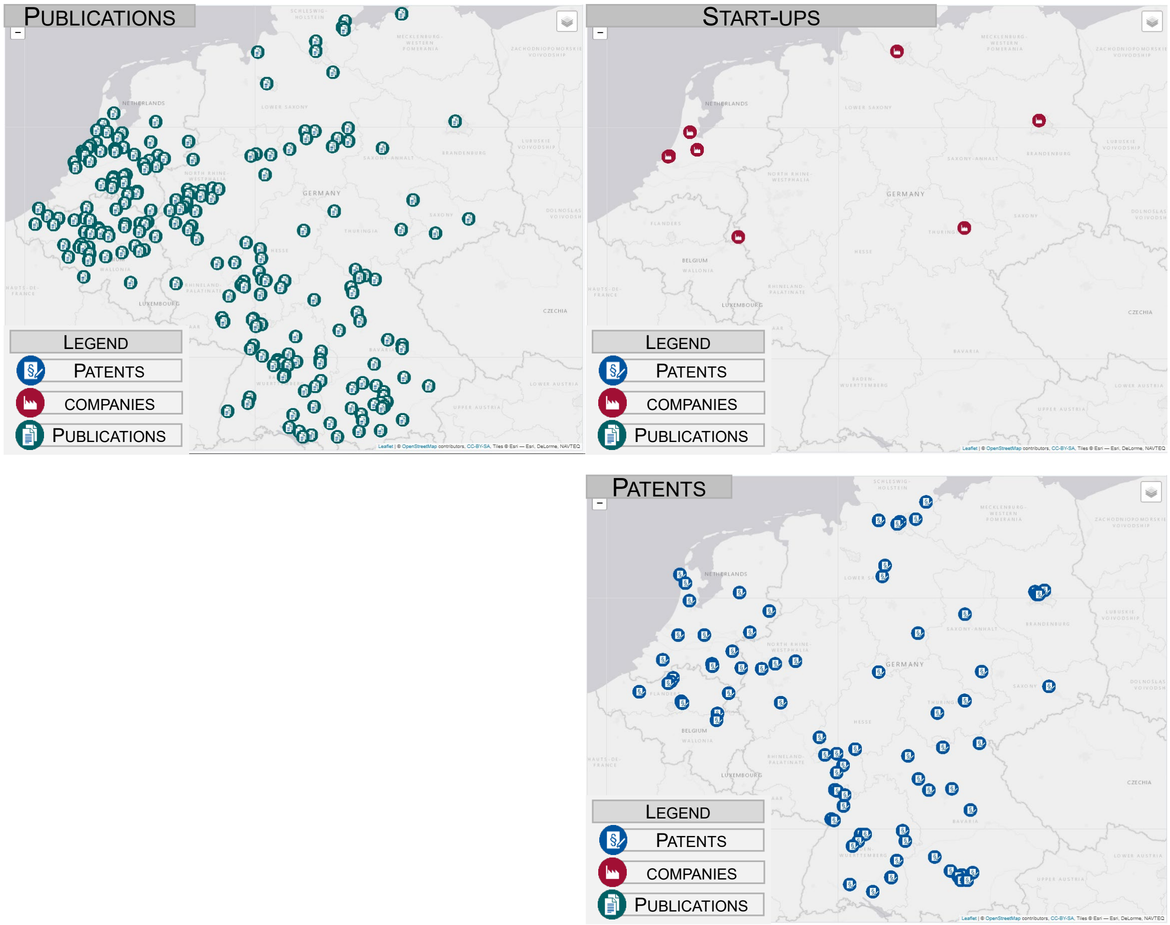
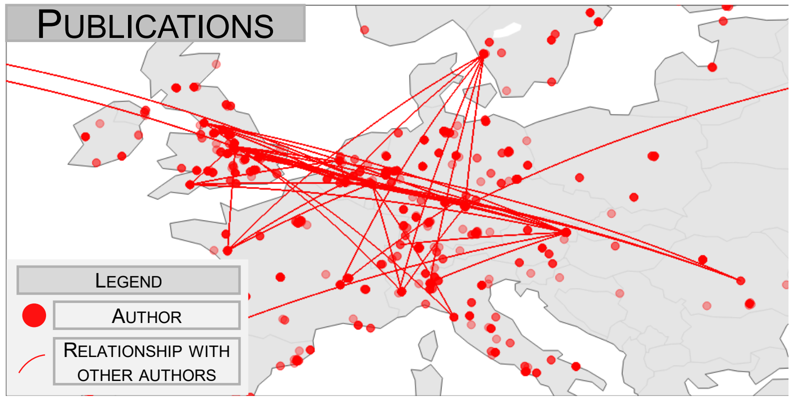


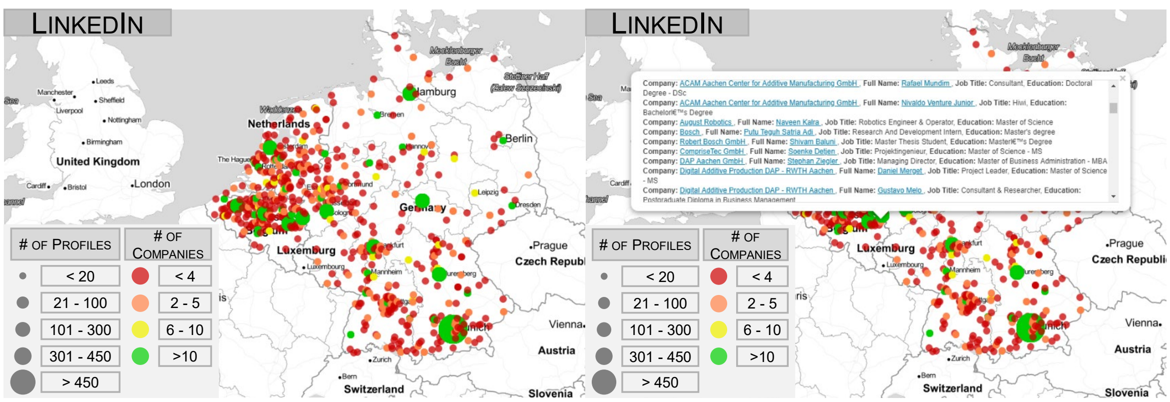
Figure 5: Exact localization of publications, patent applications and company start-ups in Germany, the Netherlands, Belgium for 2018.

Looking at the distribution of publications within Germany, agglomerations can be found in the west of NRW (Cologne, Düsseldorf and Aachen) as well as in the Stuttgart, Darmstadt and Bamberg/Erlangen area. In the Netherlands, the concentration of publications is greatest in the Amsterdam and Eindhoven area. Only the Brussels/Leuven region shows a notable publication density within Belgium. Further information (e.g. author affiliations, publication titles) is provided behind the respective publication symbol. Within Germany, these are primarily RWTH Aachen, FH Aachen, TH Köln, TU Berlin, TU Chemnitz, TU Stuttgart, TU München, LU München, Hochschule München, TU Darmstadt and the associated Fraunhofer Institutes here. In Belgium, the list is dominated by KU Leuven and in the Netherlands by TU Eindhoven and the University of Amsterdam. In Germany, on the other hand, patent applications are strongly dominated by the regions of Munich, Hamburg, Stuttgart and Berlin. In Belgium and the Netherlands, there are no significant differences between the individual locations. In contrast, business startups in the Netherlands are located exclusively in the Amsterdam region. Start-ups in Germany are located in Aachen, Hamburg, Jena and Berlin. With the exception of Jena, significant AM activities are found in all locations. Consolidating all three data sources, the area between Eindhoven (Netherlands), Leuven (Belgium) and the west of NRW (Aachen, Düsseldorf and Cologne) emerges as the strongest technology region. This is also supported by the relationship between the different authors in the EU member states for 2018 visualized in Figure 6, where it can be observed that the locations of Vienna, Aachen, Eindhoven, Leuven and the Birmingham region have increasingly published publications jointly. This is always an indicator of a scientific network. The distribution of publications over time shows the long-standing relationship between the sites.

Figure 6: Locations of the authors and their relationship to each other



In addition to publications, which can be used to increasingly identify experts in research, the systematic analysis of LinkedIn profiles is useful in identifying experts in business. In addition, further information can be obtained for evaluating market development as well as technology regions. Figure 7 shows the distribution of LinkedIn profiles in Germany, the Netherlands and Belgium. The number of profiles and companies are indexed based on marker size and color (see Figure 7, left). Each data point also includes additional information, such as total number of profiles and local companies, as well as a detailed listing of profiles (see Figure 7, right). A total of 9403 profiles from the Netherlands, Belgium and Germany are analyzed.



se: Localization of LinkedIn profiles of companies and individuals on AM in the Netherlands, Belgium, and Germany.

Compared to the information from Amadeus, the identical regions with increased activities in AM can be identified. Likewise, it can be clearly seen that the west of NRW, together with the regions of Leuven/Brussels, Amsterdam and Eindhoven, has a very high density of profiles as well as a large number of companies. This underlines the great importance of these technology regions in the AM sector. In contrast to the information obtained from Scopus and Amadeus, however, attention must be paid here to data quality, since the information input in LinkedIn is user-specific and not subject to quality assurance. This means, for example, that employees of RWTH Aachen University are located at other locations (e.g., Berlin) because the profile is not updated. Likewise, semantic designations lead to a distortion of the analysis. For example, AM and 3d printing are used synonymously and companies (e.g. Ford Werke GmbH, Ford Motors Company, Ford, Ford Motors) and locations (e.g. Aachen, Aachen region) are referred to differently. In addition, further questions can be answered with the help of this analysis. For example, the targeted acquisition of skilled workers or location decisions based on resident skilled workers. Potential customers for regional supplier networks could also be identified in a targeted manner.

Assessment of the Added Value and Conclusion

The use of the above-mentioned unconventional data sources for the AM use case can be applied for targeted knowledge gain. The predefined questions can be answered sufficiently, whereby a more detailed specification of the relevant topics in the field of AM could lead to improved knowledge gain. As an example, the publication data according to the processing of specific materials using AM can be clustered for easier identification of suitable experts. Similarly, data analysis in terms of temporal trends across regions is relevant for evaluating market development as well as competitors and should be considered in future work. Here, unconventional data sources, such as LinkedIn, play a relevant role. The analysis of LinkedIn profiles is challenging due to a lack of quality assurance. Nevertheless, this data maps changes in the market in real time and also allows for targeted competitive intelligence. The analyses developed can be directly used for application where the analyses can be profitably used. The analysis of LinkedIn profiles also can be used for lead generation. Additionally, the analysis for localizing topic-specific publications can be also for greater added value for the companies. For this, a better clustering of keywords and topics would have to be implemented. In summary, added value can be generated for companies. Another essential observation is the management of expectations for the presented analyses. These may range from very vague to very specific expectations. The following guidelines can be derived to manage expectations:

a. Posing the right question to the right data source.
b. Use more than one data source to answer the research question
c. Data sources would need to be accessible and obtainable via defined interfaces (e.g., APIs)
d. (Automatic) analysis and visualization should be used to interpret the data

Furthermore, machine learning algorithms can extend the deterministic presented approach and efficiently process larger amounts of data. Especially for social media or brand monitoring, Deep Learning algorithms can be used for text classification to transform unstructured into structured data. Here, instead of manually created rules, Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN) are used. If the amount of data is small, Support Vector Machines (SVM) can also be used.

Similarly, the Topic Modeling method can be used to explore text collections (e.g., publication data) thematically. Overall, Topic Modeling is considered to be a comparatively efficient method. One of the most common approaches is based on Latent Dirichlet Allocation (LDA), which was mainly developed by David M. Blei. This method can be used, for example, to identify the broad impact of research trends in AM or experts on predefined keywords.

Overall, machine learning algorithms provide an efficient approach to analyze large amounts of text. However, currently the use of such algorithms in the economic context is manageable and primarily focused on the finance and insurance sector. In the future, these approaches should be more widely pursued and used to unlock their full potential for business.

Acknowledgements

This research was partly funded by the Federal Ministry for Education and Research of Germany BMBF in the Project Digital Photonic Production DPP (Grant 13N15423) and partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´ s Excellence Strategy – EXC-2023 Internet of Production – 390621612. The responsibility for the content of this publication lies with the authors.

References

[1] dobney com Market Intelligence. http://www.dobney.com/market_intelligence.htm

[2] Hedin H, Hirvensalo I, Vaarnas M (2014) The Handbook of Market Intelligence. Understand, Compete and Grow in Global Markets, 2nd ed. Wiley, Hoboken

[3] Jaman, S. F. I., & Anshari, M. (2019) Facebook as marketing tools for organizations: Knowledge management analysis. In Dynamic perspectives on globalization and sustainable business in Asia, pp. 92-105. IGI Global

[4] Rheinhardt Neuhold Die besten (kostenlosen) Tools, um Website-Traffic zu messen

[5] Schindler S (2015) Strategische Planung von Technologieketten für die Produktion. Forschungsberichte IWB, vol 294. Utz, München

[6] United Nations Development Programme (2020) Education index. http://hdr.undp.org/en/indicators/103706. Accessed 08 Apr 2020

[7] Vecchione A, Brown D, Allen E et al. (2016) Tracking User Behavior with Google Analytics Events on an Academic Library Web Site. Journal of Web Librarianship 10(3): 161–175. doi: 10.1080/19322909.2016.1175330

[8] Wohlers Associates I Wohlers Report 2019. https://www.engineeringspot.de/2019/09/formnext-wohlers-investor-dinner/. Accessed 08 Apr 2020