Challenges in Implementing a University-Based Innovation Search Engine

Arman Arzani[a], Marcus Handte[b] and Pedro José Marrón[c]
University of Duisburg-Essen, Essen, Germany

Keywords: Knowledge Transfer, Founding Potential, Researcher Profiling, Innovation Identification.

Abstract: In universities, technology transfer plays an important role in the joint development and dissemination of knowledge as a product that benefits society through innovation. In order to facilitate knowledge transfer, many universities hire innovation coaches that employ a scouting process to identify faculty members and students who possess the requisite knowledge, expertise, and potential to establish startups. Since there is no systematic approach to measure the innovation potential of university members based on their academic activities, the scouting process is typically subjective and relies heavily on the experience of the innovation coaches. In this paper, we motivate the need for INSE (INnovation Search Engine) to support innovation coaches during their search for innovation potential at a university. After discussing the information needs of the scouting process, we outline a basic system architecture to support it, and we identify a number of research challenges. Our aim is to motivate vigorous research in this area by illustrating the need for novel, data-driven approaches towards effective innovation scouting and successful knowledge transfer.

1 INTRODUCTION

Technology transfer is central to the development of an iconic entrepreneurial university. Academic science has become increasingly entrepreneurial, not only through industry connections for research support or transfer of technology but also in its inner dynamic. Many universities are expanding their traditional roles beyond education and research to include knowledge transfer, which involves joint development and dissemination of knowledge as a product that benefits society through communication, experience sharing, building contacts, and innovation networks. There are various forms of technology transfer at universities, including the marketing of patents through licensing agreements, the formation of joint ventures with industrial partners as well as the creation of startups that aim to commercialize an innovative research result (Bliznets et al., 2018). While all these forms can benefit society, academic spin-offs are often considered to be particularly valuable, since successful startups can not only create new jobs and increase tax revenues but also inspire other potential founders to follow suit.

To promote knowledge transfer in general and the creation of academic startups in particular, many universities employ innovation coaches that support researchers by offering consulting services, by mediating funding possibilities facilitated through industry collaborations and investment partnerships, or by arranging workshops and startup schools. However, these activities are of a reactive nature, as they require researchers themselves, to seek out and seize these offerings. To raise the awareness and to increase the effectiveness of their knowledge transfer activities, some universities have begun to take a more proactive role. To do this, their innovation coaches are regularly performing scouting activities to identify potential innovations and innovators inside their organization. Conceptually, the innovation scouting process can be split into three stages. In the first stage, the innovation coaches identify emerging trends by gathering information about science and technology in the early stages from both formal and informal sources, including expert insights (Calvi et al., 2020). Besides from talking to experts. the coaches usually browse newsfeeds, projects, reports, scientific papers as well as patents from inside and outside the organization. Once a trend is identified, the innovation coaches find the experts within the organization. To do this, they usually rely on public information made available through the websites of the different institutes, faculties and research groups. Finally, in the last stage, the coaches prioritize the groups and indi-

[a] https://orcid.org/0009-0000-1304-9012
[b] https://orcid.org/0000-0003-4054-1306
[c] https://orcid.org/0000-0001-7233-2547

Arzani, A., Handte, M. and Marrón, P.
Challenges in Implementing a University-Based Innovation Search Engine.
DOI: 10.5220/0012263100003598
In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - Volume 1: KDIR*, pages 477-486
ISBN: 978-989-758-671-2; ISSN: 2184-3228

viduals with matching expertise and results based on their knowledge, motivation, founding potential and collaboration skills (Albers et al., 2020) in order to establish a contact with them.

Since there is no systematic approach to measure the innovation potential of university members, the scouting process is typically manual and relies heavily on the experience of the innovation coaches as well as their networking capabilities. Due to this, innovation coaches are often networking experts and exhibit an extensive network of industrial partners. However, even if the innovation coaches can use their network to easily identify enabling technologies and product gaps, the success of scouting also depends on the understanding of organizational hierarchy and its underlying individuals. This makes the systematic identification, of the experts and the ranking and evaluation of the innovation potential inside an organization a difficult challenge.

In this paper, we motivate the need for INSE (INnovation Search Engine) to support innovation coaches during their search for innovation potential. While it may not be possible to completely systematize the innovation scouting process, we believe that innovation coaches can greatly benefit from an easily accessible and up-to-date contextualized view of their organization. After discussing the information needs of the scouting process, we outline a basic system architecture to support it and we identify several challenges. Our aim is to motivate vigorous research in this area by illustrating the need for a novel, data-driven approach to support the scouting process which could further improve the effectiveness of knowledge transfer.

The remainder of the paper is organized as follows: Section 2 introduces the related work in this area; Section 3 outlines the information needs of the innovation coaches during the scouting process. Section 4 presents the INSE architecture in support of the scouting process and Section 5 describes the identified research challenges in implementing INSE for a university. Finally, in Section 6 we conclude the paper with a short summary and an outlook.

2 RELATED WORK

Academic entrepreneurship plays a pivotal role in advancing scientific progress by bridging the gap between research outcomes and real-world applications through spin-offs and has been the focus of many studies in the research community (Rippa and Secundo, 2019; Schultz, 2021; Kalinowski, 2016). While (Rippa and Secundo, 2019) indicates and addresses the lack of in theoretical development in this domain by underlining the necessity for more inclusive studies that encompass technological, economic, and social dimensions of academic entrepreneurship, (Kalinowski, 2016) describes its weaknesses in existent research transfer systems by an analysis of scouting implementation experiences in Polish universities. Thereby, technology scouting is described as a strategic response to address the limitations of the commercialization system by involving a systematic approach to gather information in the realm of science and technology and seeking bidirectional exploration for novel opportunities in specific technological fields. Furthermore, technology scouting and its quality are also emphasized by an analysis of academic spin-offs of University of Potsdam (Schultz, 2021) as a beneficial factor that should be combined with other systematic scouting activities in order to create a sustainable raise in academic entrepreneurship.

Technology scouting is a strategic approach that involves identifying and incorporating innovations, such as those from startups, university intellectual properties and acknowledges the increasing complexity of products and services, recognizing that companies and universities can not solely rely on internal efforts to innovate effectively (Wang and Quan, 2021). Therefore, many works include technology/innovation scouting frameworks to promote startups and innovation either in universities or companies through tools for technology/knowledge transfer such as technology radars(Rohrbeck et al., 2006; Golovatchev and Budde, 2010; Desruelle and Nepelski, 2017; Berndt and Mietzner, 2021).

For instance (Rohrbeck et al., 2006) discusses the use of technology scouting in Deutsche Telekom Labs, where they employ a technology radar approach to enhance traditional newsletter-based scouting. This method facilitates communication between experts and scouts, improving the scouting process through networking efforts. It involves an international scouting network that analyzes trends, publications, and patents to identify potential technologies. The scouts manually assess these technologies, and an expert panel evaluates and ranks them, generating a report as the output. Additionally, as an extension (Golovatchev and Budde, 2010) introduces innovation radars, which focus on market aspects of technology and rely on reactive expert inquiries for technology identification and evaluation. Prior research primarily emphasizes accelerating innovation within corporate settings through technology and innovation radars. While their methodologies can also be adapted for application within university contexts to aid future startups, the authors of (Berndt and Mietzner, 2021) focus on digitizing these radars, considering them as tools to facilitate knowledge and technology transfer in academic setting. They showcase their outcomes through a web-based technology radar, serving as an online collaborative tool. Their argument centers on the absence of systematically explorative insight for technology scouts and highlights how employing these tools significantly simplifies the complexity of their evaluations.

There have been studies examining the role of patents in advancing academic entrepreneurship. These investigations have shown that the incorporation of knowledge transferred from the parent university and academic founders through patents has a noticeable impact on the success of Academic Spin-Offs (Ferri et al., 2019). In fact some works put their focus on utilizing patent analysis to support the technology/innovation scouting process by identifying technology trends or new business opportunities (An et al., 2018; Lee and Lee, 2017; Chen et al., 2015). The work of (An et al., 2018; Lee and Lee, 2017) uses text mining to derive keywords on the technology hotspots from the patent content, while (Chen et al., 2015) combine this technique with piecewise linear representation of patent publications throughout time to explore their hidden technology trend patterns.

The aforementioned approaches prove the applicability and importance of technology and innovation scouting for the identification of academic startups through systematic frameworks and patents. However, these approaches lack the systematic data-driven approach, as some rely heavily on manual human interactions, or in the case of patent-oriented technology scouting only consider patents as the main drivers of innovation. Innovation scouting for the creation of academic spinoffs should consider various data sources aside from patents such as publications or media along with data-driven processes involving the identification of university expertise. Despite the theoretical depth of academic entrepreneurship in related work, there is clearly a gap for implementing a data-driven solution for identifying the innovation potential at universities.

3 INFORMATION NEEDS

As outlined in the introduction, innovation scouting can be split into three stages, namely the identification of technology trends, the search for expertise and the assessment of innovation potential. Each stage involves the collection of specific data that is eventually used as the basis for decisions. Thus, scouting not only encompasses data gathering but also requires an interpretation that is specific to each stage. In the following, we discuss the goals of each stage and derive the resulting information needs. Although, our description follows a top-down approach that allows an innovation coach to identify candidates that match a particular trend, it is noteworthy to point out that, in practice, an innovation coach could also follow a bottom-up approach that tries to determine how well a certain researcher or research result matches a trend.

3.1 Identification of Technology Trends

Knowledge about the current technology trends is an important input for innovation scouting. Although, some innovations may establish new trends by themselves, it is often easier to find partners or acquire seed funding, if the intended innovation domain matches or extends an existing trend. Thus, the goal of this stage is to monitor and understand the current and emerging technology trends within different markets and industries. Thereby, the coach gains insights into what the market demands, where opportunities lie, and which directions innovation might take.

Identifying the technology trends involves research and analysis of various factors such as consumer preferences, cultural influences, and technical advancements, to name a few. This means that innovation coaches must follow the latest advancements in the target sector. To do this, they can tap into regular news sources (e.g., to cover consumer preferences or cultural influences) as well as publications and patents (e.g., to cover technological developments). In addition, they must relate the advancements with past developments to judge their significance. Since this is a complex undertaking that requires experience, innovation coaches often have an extensive industry background and are in active exchange with their network of collaboration partners.

3.2 Search for Expertise

Given the knowledge about the technical trends, innovation scouting must match the trends with the expertise inside the organization. For a university, this could be a certain faculty, an institute, a research group or a particular researcher that have theoretical and practical knowledge of the subject. To determine the expertise of these entities, an innovation coach must screen their research activities and results. The latter can be extracted by analyzing the publications or patents created within the organization. For the former, it is possible to look at grants awarded to the different entities in the organization and the associated research projects in which they are involved.

Besides from determining which entities exhibit expertise in a particular domain, it may also be necessary to judge whether the level and type of expertise can serve as a basis for technology transfer. To clarify this, consider that in a large organization such as a typical university, the different entities are often not focused on a single problem but are working on a broad range of topics. Thus, although an entity is working in a particular domain, it might not be its primary focus. Similarly, while some entities might be addressing problems that exhibit an immediate relevance to industry, other entities might be focusing on addressing basic research problems that will only become relevant for technology transfer in the future.

3.3 Assessment of Innovation Potential

After identifying a technology trend and finding the relevant expertise within the organization, the goal of the last step is to determine the set of persons that should be contacted by the innovation coach. Thus, given a set of entities exhibiting the targeted type and level of expertise, it is necessary to identify the most likely persons to participate in or drive the creation of a startup. Towards this end, an innovation coach must first identify the relevant persons associated with an entity and then assesses their background, skills, knowledge, and abilities in order to prioritize them.

To identify the persons associated with an entity, an innovation coach can analyze the organization structure. To create a ranking, the coach can then compare the background information of different researchers within the organization. This might include, for example, the CVs of the researchers, their publication lists or their patent portfolios. In addition, the coach might want to compare the grants and projects or look at the collaboration networks of the researchers.

4 INNOVATION SEARCH ENGINE

As described previously, each stage of the scouting process requires both data gathering and interpretation. For data gathering, innovation coaches must sift through a broad range of information sources.

Examples include newsfeeds, research projects, reports, grants, scientific literature, and patents, both internal and external to the organization. Depending on the stage, they must either organize the data over time, e.g., to determine trends, or they must link the information with the different entities within their organization, e.g., to determine the level and type of expertise of a particular entity inside the organization. For interpretation, innovation coaches must aggregate the data from different sources and analyze the results. This includes the comparison of aggregations, e.g., to rank technology trends or to prioritize persons that should be contacted. At the present time, there is no thorough system-support for innovation scouting at universities. As a result, data gathering is a complicated and labor-intensive task. For information that is directly available via the World Wide Web, innovation coaches can use search engines such as Bing and Google as their entry point. For restricted information, such as (non-open access) publications or internal information about the organization, the coaches must individually access a potentially large range of information systems. While most of these systems typically exhibit some way of searching for a particular piece of information, they usually lack the necessary aggregation capabilities that are required for innovation scouting. Thus, besides from searching, the data gathering usually involves manual aggregation within and across different data sources.

A similar argument can be made about data interpretation. Given that there is no systematic approach to measure the innovation potential, the data interpretation required to identify trends or to prioritize individuals usually relies heavily on the experience and networking capabilities of the innovation coaches. As a result, the decision-making during innovation scouting is a rather subjective process. While it may not be necessary to completely systematize the scouting process, we believe that it can be greatly simplified by providing adequate system-support.

To this end, we started the development of INSE (INnovation Search Engine), an application-specific search engine that aims to provide thorough support for all three stages of innovation scouting. INSE automatically gathers information from relevant data sources such as newsfeeds, research projects, reports, scientific literature, and patents, both internal and external to the organization. Then INSE stores and analyzes the data. Thereby, it extracts relevant information and links the different pieces of information across different data sources. Using the linked information, INSE offers data aggregation, search, and browsing functionalities to the innovation coaches. On top of this, INSE offers a range of metrics that support the three stages of innovation scouting. The goal thereby is not to replace the (expert) opinions of the innovation coaches with metrics, but to provide them with a compact, up-to-date, and contextualized view of the organization. Figure 1 shows a high-level architectural overview of INSE. As part of an ongoing research project, we are in the process of implementing the different components of INSE in order to support the innovation coaches of our university. In the following, we specify each component and describe their corresponding inputs and outputs.

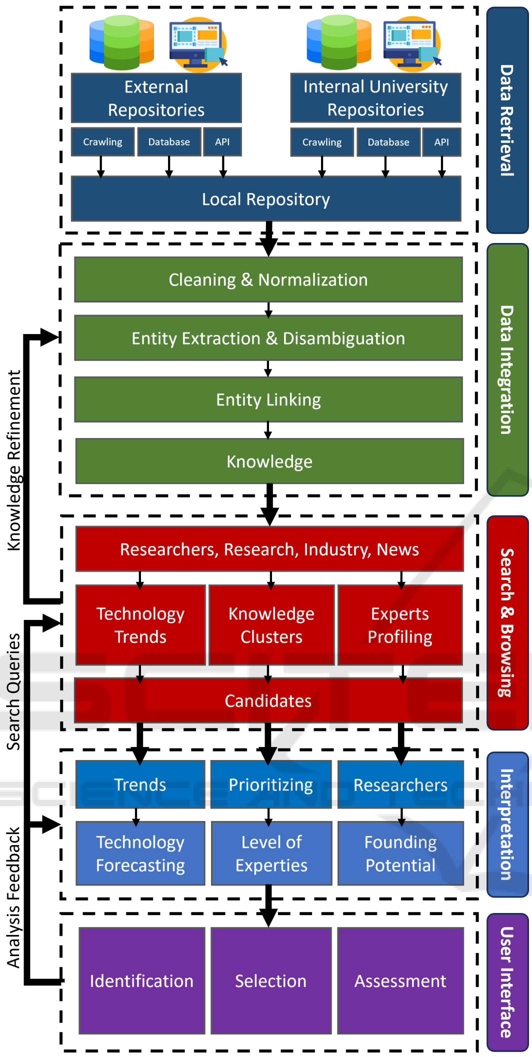

Figure 1: Architecture of Innovation Search Engine.

4.1 Data Retrieval

The data retrieval component is responsible for the gathering of data from various data sources. This component aims to minimize the need for manual data gathering by storing relevant data from multiple information systems in a single data repository. The data retrieval process occurs through direct access to the databases, APIs, as well as web crawling. For instance, while a university maintains a library database for the scientific publications of a certain researcher, the information of their patents, participated research grants, faculty, research group or institute might not either be stored in a single database, or it is often scattered among many databases. Therefore, the data retrieval merely depends on the type of data source.

The target data sources include (1) internal data sources of the targeted university; (2) external data sources such as scientific libraries for research papers, patent repositories, industry reports and news portals. While the internal data sources reflect the targeted university's relevant information for the innovation coaches, the external data sources provide contextual information as basis for comparison.

For internal data sources, INSE handles data gathering via integrated connectors that have been developed to support different input formats. In addition, INSE crawls the public website of the university to gather information about the organizational structure and the research activities. To do this, INSE starts from a set of URLs given as an input parameter and traverses the graph of pages by following the links contained in pages. Since the pages may link to other pages outside the university, the traversal can optionally be restricted by a list of target domains. During the traversal of the graph, INSE automatically extracts the web pages of the organizational hierarchy and its underlying entities along with the attached metadata. These entities may include information on institutes, faculties, chairs, researchers research papers, funding projects, news, and patents.

For external data sources, INSE gathers data either via the APIs offered by the data provider or via web crawling. Some of these sources that we have added to INSE include libraries of patents such as European patent office (EPO), United States Patent and Trademark Office (USPTO), scientific portals (Scopus, Web of Science, DBLP), industry reports (Crunchbase) and news (Google Trends).

The actual data gathering strategy depends on the source. For instance, in case of patent libraries, EPO offers patent records that can be directly accessed through an API, while the USPTO provides bulk records that can be crawled. After gathering the data from internal and external repositories, INSE stores, indexes the data in a local repository for later search and browsing.

4.2 Data Integration

The data integration component has two main goals, namely entity extraction and linking of relevant information across different data sources. The gathered data in the previous step includes structured and unstructured information. While extracted information from available data repositories (e.g., databases, APIs) deliver organized fields and data formats, this is not the case for the uncurated crawled data.

INSE employs Natural Language Processing (NLP) pipelines to preprocess the crawled data. Thereby, INSE performs data cleaning of stored unstructured text data through the removal of language-specific stop words and Unicode characters, as well as lemmatization and stemming. This step cleans and normalizes the contents of the local repository. After cleaning and normalization, INSE aims to extract the entities within the local repository. INSE utilizes the Entity Recognition techniques (NER) to identify and mine entities such as faculties, institutes, groups, researchers, papers, patents, and news. Given that the identified entities hold references to their content, INSE utilizes their content to disambiguate them to increase the quality of entities. For example. entities such as Thomas Müller, Müller Thomas, Müller Th. should point to the same entity depending on their content, faculty, or research group.

After the identification of the entities, each entity needs to be further described by the corresponding metadata. For instance, in the case of a researcher, information of their patents, publications and mentions in news articles or industry reports should be linked with and integrated into their profile. INSE realizes this through linking of entities across gathered internal and external sources that are stored in the local repository. As its final output, the integration component produces a knowledge repository, holding the extracted entities and their linkage to other entities.

4.3 Search and Browsing

The search and browsing component serves as an entry point to the innovation search engine. This component offers active-search and explorative-search interfaces. It handles data aggregation, search, and browsing functionalities for the innovation coaches.

By using the active-search interface, the innovation coaches query and navigate through extracted entities and technology trends in industry, research, and news articles. Thereby, INSE can provide an in-depth overview of expert profiles based on entity categories or distinct technologies through dashboard visualizations for the user interface. Search queries offer insight of the organizational information such as faculty or research group of an individual, as well as the linked scientometrics and bibliometrics about the research activities and their results. Moreover, the dashboards present information on the collaborations between researchers, research/industry project partners, that may unveil knowledge clusters and potential collaborations. The active-search interface also taps into the gathered data from external sources and demonstrates the comparative analysis of technology trends to both inside and outside the university.

The explorative-search interface offers the same set of functionalities in an exploratory fashion. This is helpful for innovation coaches who require a general overview of the university entities and their activities or may not search for a specific technology area or individual. The explorative dashboards show the active technology trends to both inside and outside the university and its researchers. Also, the explorative dashboard also aggregates data and visualizes pointers to the most active entities and their expertise in and outer university, based on number of papers, patents, news mentions.

Furthermore, INSE offers user interactions in the search component to tag invalid or missing links in order to give feedback to the data integration component regarding the quality of entities and their linkage, so that they can be further adjusted. Finally, the innovation coaches can use the search interfaces, to generate and bookmark a set of candidates, which serve as the output of the search component. Depending on the stage of the scouting process, these candidates consist of selected technology trends, researcher(s), groups or collaboration clusters.

4.4 Interpretation

Given a set of experts and technologies, the goal of the interpretation component is to provide metrics on the technology trends and individual(s) by measuring the innovation potential of a particular technology or individual(s). This is done in three steps that match the stages of the scouting process described in Section 3. (1) modelling technology lifecycles; (2) prioritizing experts; (3) modelling innovation potential. To tackle the first step, INSE aims to generate the technology lifecycle to measure the technology's readiness and maturity. Thereby, INSE leverages the stored local repository to analyze patterns of the historical data on past emerging technologies and breakthroughs through patents and publications as well as hype from media and news. As a result, INSE classifies the trendiness of a technology based on a mathematical model that tries to capture the various phases of the technology lifecycle.

Then in the second step, INSE prioritizes the experts based on the level of expertise for each of the selected technologies, by taking into account the expert's conference and journal impact score as well as citation metrics for their publications and patents. This provides a ranked list of experts that are active in a particular domain.

As a last step, since being an expert in a specific field does not necessarily indicate founding potential, INSE leverages the stored local repository to analyze patterns of the historical data on academic founders. Thereby, INSE utilizes a founder database extracted from technology portals such as Crunchbase and its linkage to the entities in the local repository to compile a list of founders and non-founders (Arzani et al., 2023). By generating a dataset of features for founders and non-founders, patterns of founding potential are modeled with a discriminative machine learning algorithm such as Random Forests or a binary classification deep learning model. Example for a set of features could include number of patents, publications or linked patents to industry or funded projects as well as citation metrics. This allows INSE to measure the innovation potential for each of the individuals in the prioritized list of experts.

The output of the interpretation component consists of the lifecycles of the candidate technologies and a list of prioritized experts, as well as the list of high potential startup founders.

4.5 User Interface

The user interface builds up on the output of interpretation component and consist of the well-timed technologies, ranked experts and high potential innovators. This component also describes the main three tasks of the innovation scouting process, that are realized through the previous components. The UI generates reports for the identified technologies and the selected candidates in the search component. Furthermore, the component offers interactive dashboards and visualization to the innovation coaches via web interfaces on the phases of technologies and the profiles of the selected candidates. The UI also visualizes bar charts on the ranked experts, their assessment metrics as well as their estimated founding potential.

Additionally, the web interface explains a roadmap of made decisions in the scouting activities to the innovation coaches that led to the presented results. This generates an overview of the selected technology trends, researchers, and their collaboration network in the search component. Also, the UI shows the transition to the interpretation component by including a description of applied models and their parameters to offer explainability and future model improvements based on the coach's feedback. Ultimately the innovation coach takes their own initiative and decides whether they want to establish contact for a startup opportunity.

5 CHALLENGES

During our ongoing work on implementing the INSE architecture described above to support innovation scouting at our university, we identified three main challenges. The first challenge is **data acquisition** that describes the difficulties of data gathering in the context of innovation scouting. The second challenge refers to **data integration** that describes the issues resulting from the need to link information from multiple data repositories into a coherent knowledge repository. Lastly, the challenge of **data interpretation** underscores the analysis and data modelling that aggregates the available data into assessment metrics and actionable intelligence. In the following, we provide discussion on each of these challenges.

5.1 Data Acquisition

The scouting process requires access to multiple data sources that include newsfeeds, research projects, reports, grants, scientific literature, and patents, both internal and external to the university, whereas the external data sources determine a university's expertise and role (e.g., leader, follower) in a trending topic.

Internally, the challenge arises from the absence of a central repository at the university housing all the required data. As a result, the required data is often spread across several systems that are operated by different entities inside the university. Thereby, it is noteworthy to mention that despite being present in some system, some data might not be accessible or might even be deliberately blocked due to legal or privacy reasons. Examples of systems that are maintaining relevant data include financial systems (e.g., for grant details or employee lists), library systems (e.g., for publications) and web pages (e.g., for research activities). For the latter faculties or research groups often maintain and update their own content management systems that may not only exhibit different structures but also present information such as the research focus or ongoing research activities at different levels of detail. For example, some groups might include information on the staff, publications, and projects but not necessarily bibliometrics on publications or even patents or news. This means that the missing information must be compensated through internal sources (e.g., university-wide news feeds) or external data sources (e.g., patent offices). Thus, maximizing the data availability can become quite challenging, since custom database connectors and crawlers might have to be specifically adjusted to each data source available inside the university.

The challenge of data acquisition is mirrored externally, where relevant publication and patent sources are dispersed among various publishers or trademark offices. Data such as patents and publications are also available in scientific libraries behind paid API calls that pose query limitations.

5.2 Data Integration

After acquiring the data from multiple sources, the data must be linked to support the innovation coaches by generating information about researchers' profiles, faculties, and groups. This involves the integration of patents, publications, and projects with specific individuals and organizational units. This process necessitates creating connections between the gathered data sources and the internal data infrastructure of the university. Before linking, these entities first have to be identified and disambiguated throughout all the available data sources. The disambiguation is a common challenge in cases where there are researcher entities with similar names or affiliations. This also includes distinguishing between multiple researchers with the same name or identifying variations in naming conventions across different publications and databases. To this end, general efforts i.e., use of an Open Researcher Contributor Identification (ORCID) or other advanced techniques, such as name disambiguation algorithms, network analysis, and affiliation disambiguation are necessary to ensure accurate and distinct researcher profiles.

5.3 Data Interpretation

After retrieval and integration of multiple data sources into a single repository and generating the linkage between the entities, INSE needs to interpret the patterns of data via modelling in order to extract knowledge and deliver actionable intelligence to the innovation coaches. The challenges in data interpretation encompass three aspects for an effective scouting process which we discuss in detail. The first challenge refers to **technology forecasting** that signifies the study of technology trends through the analysis of technology life cycles. The challenge of **level of expertise** revolves around the ranking of the experts based on their scientific profile that helps the innovation coaches to filter out the most suitable candidates. Lastly, the challenge of **founding potential** underscores the evaluation of founding potential based on academic background.

5.3.1 Technology Forecasting

Technology forecasting and technology trends through a combination of news articles, patents and scientific publications has been proven effective and has taken the attention of many researchers (Chen and Han, 2019; Asooja et al., 2016; Winnink et al., 2019). Indeed, a clear picture of technology life cycles is required to measure the impact of emerging technologies that facilitates the innovation coaches' ability to anticipate technological breakthroughs for upcoming startups. The challenge here is to define the maturity and readiness of a technology. Thereby, a novel baseline for the maturity of technology should be developed based on R&D activities such as scientific literature, patents as well as social media. Defining this baseline provides a better understanding of technology life cycles. One of the ways to discuss the phases of a technology's is the Gartner hype cycles (Chen and Han, 2019). Many scholars investigated the methodology of the Gartner hype cycle in their work (Dedehayir and Steinert, 2016; Van Lente et al., 2013). While the authors of (Dedehayir and Steinert, 2016) provide evidence that the Gartner model is a combination of hype level through media with the business maturity curve also known as S-curve, (Van Lente et al., 2013) argue for the lack of mathematical foundation of the Gartner model. Furthermore, Gartner hype cycle is based on quantitative analysis, as there is no single measure through surveys, evidence and forecasts, meaning that the cycle involves expert judgment, which can be quite bias. There are also reports of discordance between the generated Gartner models for some years and the actual trend of technologies based on scientific literature and patents (Chen and Han, 2019). Considering this challenge provides an opportunity for further in depth analysis based on bibliometrics and scientometrics to improve the explainability of technology hype cycle models.

5.3.2 Level of Expertise

After determining the technology area and identifying the knowledge experts or faculties at a university, another factor of successful knowledge transfer is the level of research done by the researchers. This includes the weight of parameters such as conference and journal rankings, as well as impact factors that are dependent on the type of disciplines.

The evaluation of research data through conference and journal rankings, along with impact factors, carries varying weight across different fields of study. Depending on the discipline, the prestige of known conferences and journals can directly correlate with technological advancements. Also, the impact factor of a conference or journal also serves as a potential metric for academic success based on the main discipline. These metrics vary across domains, since in certain fields, innovative ideas and concepts may emerge from less conventional channels. For example, while journals in the medical field often have relatively high impact factors, the same can be said about international conferences in the engineering discipline. Finding a valuable metric through evaluation of venues and their impact factors provides a list of high ranking experts inside a university.

5.3.3 Founding Potential

Universities increasingly encourage the involvement of their academics in the transfer of knowledge to the marketplace through spin-off activities (Siegel and Wright, 2015; González-Pernía et al., 2013). Study based on a comparative analysis between Japanese and US startup scene, indicates that the R&D activities serve as an important factor in entrepreneurship activities (Kegel, 2016). Indeed, various researchers (Farre-Mensa et al., 2016; Cagnani et al., 2022; Conti et al., 2013; Helmers and Rogers, 2011) also argue a direct correlation specifically between patent activities as one of the main drivers of innovation and founding startups. Therefore, considering the universities as a vast R&D network, a data-driven scouting process can model the patterns of technology transfer that come from patenting, scientific publications and research grants.

Whereas previous studies for innovation identification, mainly depend on surveys and empirical results from individual institutions (Chung, 2023; Montebruno et al., 2020; Rivera-Kempis et al., 2021; Sabahi and Parast, 2020), there is a gap that has to take academic activities into consideration for innovation scouting at universities. The challenge here is to tap into the academic information of researchers to identify existing founding potential. Thereby, there is a need of list of previous founders which requires access to specialized startup data sources. Also, the startup data sources do not necessarily hold academic information of their founders, therefore the founder profiles have to be linked with other academic data sources. By building a dataset of previous founders and their academic behavior, the founding potential of future founders can be measured and recommended to the innovation coaches. However, there is no data source that links the founder information to their academic background at the moment.

6 CONCLUSIONS

Technology transfer plays an important role in universities for the joint development and dissemination of knowledge as a product that benefits society through innovation. In order to facilitate knowledge transfer, many universities hire innovation coaches that employ a scouting process to identify faculty members and students who possess the requisite knowledge, expertise, and potential to establish startups. Since there is no systematic approach to measure the innovation potential of university members based on their academic activities, the scouting process is typically subjective and relies heavily on the experience of the innovation coaches.

In this paper, we motivate the need for an Innovation Search Engine as an integrated solution to support the scouting process. To do this, we described the information needs of innovation coaches, and we presented an architecture to cover them. Based on our ongoing work on implementing the architecture, we identified three challenges in order to motivate vigorous research in this area. Our goal for future research is to focus on data interpretation by building and evaluating models for technology hype cycles, as well as ranking and innovation potential of researchers.

ACKNOWLEDGEMENTS

This work has been funded by GUIDE REGIO which aims to improve the ability of the science support center of the University of Duisburg-Essen in the identification, qualification, and incubation of innovation potentials.

REFERENCES

Albers, A., Hahn, C., Niever, M., Heimicke, J., Marthaler, F., Spadinger, M., et al. (2020). Forcing creativity in agile innovation processes through asd-innovation coaching. In *Proceedings of the Sixth International Conference on Design Creativity (ICDC 2020)*, pages 231–238.

An, J., Kim, K., Mortara, L., and Lee, S. (2018). Deriving technology intelligence from patents: Preposition-based semantic analysis. *Journal of Informetrics*, 12(1):217–236.

Arzani, A., Handte, M., Zella, M., and Marrón, P. J. (2023). Discovering potential founders based on academic background.

Asooja, K., Bordea, G., Vulcu, G., and Buitelaar, P. (2016). Forecasting emerging trends from scientific literature. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 417–420.

Berndt, M. and Mietzner, D. (2021). Facilitating knowledge and technology transfer via a technology radar as an open and collaborative tool. *New Perspectives in*

Technology Transfer: Theories, Concepts, and Practices in an Age of Complexity, pages 207–230.

Bliznets, I. A., Kartskhiya, A. A., and Smirnov, M. G. (2018). Technology transfer in digital era: legal environment. *Journal of History Culture and Art Research*, 7(1):354–363.

Cagnani, G. R., da Costa Oliveira, T., Mattioli, I. A., Sedenho, G. C., Castro, K. P., and Crespilho, F. N. (2022). From research to market: correlation between publications, patent filings, and investments in development and production of technological innovations in biosensors. *Analytical and Bioanalytical Chemistry*, pages 1–9.

Calvi, R., Pihlajamaa, M., and Servajean-Hilst, R. (2020). Innovation scouting: a new challenge for the purchasing function. *The nature of purchasing: Insights from research and practice*, pages 295–313.

Chen, H., Zhang, G., Zhu, D., and Lu, J. (2015). A patent time series processing component for technology intelligence by trend identification functionality. *Neural Computing and applications*, 26:345–353.

Chen, X. and Han, T. (2019). Disruptive technology forecasting based on gartner hype cycle. In *2019 IEEE technology & engineering management conference (TEMSCON)*, pages 1–6. IEEE.

Chung, D. (2023). Machine learning for predictive model in entrepreneurship research: predicting entrepreneurial action. *Small Enterprise Research*, pages 1–18.

Conti, A., Thursby, J., and Thursby, M. (2013). Patents as signals for startup financing. *The Journal of Industrial Economics*, 61(3):592–622.

Dedehayir, O. and Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, 108:28–41.

Desruelle, P. and Nepelski, D. (2017). The'innovation radar': A new policy tool to support innovation management. *Available at SSRN 2944104*.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2016). The bright side of patents. Technical report, National Bureau of Economic Research.

Ferri, S., Fiorentino, R., Parmentola, A., and Sapio, A. (2019). Patenting or not? the dilemma of academic spin-off founders. *Business Process Management Journal*, 25(1):84–103.

Golovatchev, J. and Budde, O. (2010). Technology and innovation radar - effective instruments for the development of a sustainable innovation strategy. In *2010 IEEE International Conference on Management of Innovation and Technology*, pages 760–764.

González-Pernía, J. L., Kuechle, G., and Peña-Legazkue, I. (2013). An assessment of the determinants of university technology transfer. *Economic Development Quarterly*, 27(1):6–17.

Helmers, C. and Rogers, M. (2011). Does patenting help high-tech start-ups? *Research Policy*, 40(7):1016–1027.

Kalinowski, B. (2016). Increasing the potential for commercialisation of innovation and research results within polish universities. *Modern Management Review*, 21(23):2.

Kegel, P. (2016). A comparison of startup entrepreneurial activity between the united states and japan. *Journal of Management Policy & Practice*, 17(1).

Lee, M. and Lee, S. (2017). Identifying new business opportunities from competitor intelligence: An integrated use of patent and trademark databases. *Technological Forecasting and Social Change*, 119:170–183.

Montebruno, P., Bennett, R. J., Smith, H., and Van Lieshout, C. (2020). Machine learning classification of entrepreneurs in british historical census data. *Information Processing & Management*, 57(3):102210.

Rippa, P. and Secundo, G. (2019). Digital academic entrepreneurship: The potential of digital technologies on academic entrepreneurship. *Technological Forecasting and Social Change*, 146:900–911.

Rivera-Kempis, C., Valera, L., and Sastre-Castillo, M. A. (2021). Entrepreneurial competence: Using machine learning to classify entrepreneurs. *Sustainability*, 13(15):8252.

Rohrbeck, R., Heuer, J., and Arnold, H. (2006). The technology radar - an instrument of technology intelligence and innovation strategy. In *2006 IEEE International Conference on Management of Innovation and Technology*, volume 2, pages 978–983.

Sabahi, S. and Parast, M. M. (2020). The impact of entrepreneurship orientation on project performance: A machine learning approach. *International Journal of Production Economics*, 226:107621.

Schultz, C. (2021). Does technology scouting impact spinout generation? an action research study in the context of an entrepreneurial university. *New Perspectives in Technology Transfer: Theories, Concepts, and Practices in an Age of Complexity*, pages 107–128.

Siegel, D. S. and Wright, M. (2015). Academic entrepreneurship: time for a rethink? *British journal of management*, 26(4):582–595.

Van Lente, H., Spitters, C., and Peine, A. (2013). Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change*, 80(8):1615–1628.

Wang, C.-H. and Quan, X. I. (2021). The role of external technology scouting in inbound open innovation generation: Evidence from high-technology industries. *IEEE Transactions on Engineering Management*, 68(6):1558–1569.

Winnink, J., Tijssen, R. J., and Van Raan, A. (2019). Searching for new breakthroughs in science: How effective are computerised detection algorithms? *Technological Forecasting and Social Change*, 146:673–686.