Innovations and Emerging Technologies: A Study of the Italian Intellectual Property Knowledge Database

Annamaria Demarinis Loiotile[1,2][a], Francesco De Nicolò[1,2][b], Alfonso Monaco[1][c], Sabina Tangaro[3][d], Shiva Loccisano[4], Giuseppe Conti[5], Adriana Agrimi[6], Nicola Amoroso[7][e] and Roberto Bellotti[1][f]

[1]Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro, Bari, Italy
[2]Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Bari, Italy
[3]Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, Bari, Italy
[4]Behold srl, Bologna, Italy
[5]Netval – Network per la Valorizzazione della Ricerca Universitaria, Lecco, Italy
[6]Direzione Ricerca, Terza Missione e Internazionalizzazione, Università degli Studi di Bari Aldo Moro, Bari, Italy
[7]Dipartimento di Farmacia-Scienze del Farmaco, Università degli Studi di Bari Aldo Moro, Bari, Italy

Keywords: Natural Language Processing, Italian Patents, Intellectual Property Analytics, Clustering, Technology Transfer, Knowledge Database, Healthcare 4.0.

Abstract: A great mine of innovation is represented by the excellence of the scientific know-how of the Italian universities and research centers. But very often university patents remain unvalued and unexploited, in the so-called "Valley of death". In the framework of Intellectual Property Analytics and Patent Informatics, this paper analyses the Italian patent database "Knowledge Share" and its proposed classifications (10 technological areas). By means of Natural Language Processing (NLP) techniques, we examined 1694 patents from 89 Italian Research Institutions and a cluster analysis revealed the existence of 8 homogeneous clusters instead of the 10 proposed by the platform. Thus, our findings suggest the presence of possible inhomogeneities within the traditional classifications, probably due to the emergence of novel technologies or cross-domain areas, e.g., Healthcare 4.0; moreover, these clusters could lead to better performance in terms of offer/demand matching for the platform users.

1 INTRODUCTION

The so called Fourth industrial revolution is generally identified with: exponential evolutions, integration of technologies and holistic system impact across society, industry, and countries (Schwab, 2017). The integration of Industry 4.0 technologies within society is pivotal for resolving many challenges that the world and its population are currently facing (Bartoloni et al., 2022). In this multidisciplinary and complex context, the technology transfer plays a key role for the adsorption and dissemination of technologies, resources, and knowledge to transform each invention into tangible and useful innovation.

Furthermore, the increased data availability represents an opportunity to better support decision-making processes and introduce disruptive technologies (Baglieri & Cesaroni, 2013; Aristodemou & Tietze, 2018). For this aim, companies seek effective strategies for developing

[a] https://orcid.org/0000-0003-3503-9952
[b] https://orcid.org/0000-0003-3036-8108
[c] https://orcid.org/0000-0002-5968-8642
[d] https://orcid.org/0000-0002-1372-3916
[e] https://orcid.org/0000-0003-0211-0783
[f] https://orcid.org/0000-0003-3198-2708

Demarinis Loiotile, A., De Nicolò, F., Monaco, A., Tangaro, S., Loccisano, S., Conti, G., Agrimi, A., Amoroso, N. and Bellotti, R.
Innovations and Emerging Technologies: A Study of the Italian Intellectual Property Knowledge Database.
DOI: 10.5220/0011627000003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 2*, pages 75-86
ISBN: 978-989-758-623-1; ISSN: 2184-433X

and applying technologies while observing the constraints of time, budget, and without infringing any third-party Intellectual Property (IP).

The analysis of patents can meet these companies needs because intellectual property (IP) related documents represent valuable and recognized sources of technological and legal knowledge (Aristodemou et al., 2017). As a matter of fact, different documents, such as patents, trademarks and other IP registered in national and international IP systems, contain important research results which are of great value for industry, legal researchers, and policy advocates in science and technology R&D (Trappey, Trappey & Chung, 2017).

Intangible assets such as R&D, inventions, artistic and cultural creations, brands, software, know-how, business processes and data "are the cornerstones of today's knowledge economy" (Press release EU, 2020).

As reported by the European Commission, the volume of annual investment in "intellectual property assets" has increased by 87 percent in the EU over the past two decades, in contrast to the volume of tangible (non-residential) investment. Thus, industries that make intensive use of intellectual property play an essential role in the EU economy and provide good and sustainable jobs for society (Press release EU, 2020). Seemingly, according to a study conducted by the Ponemon Institute LLC looking at the S&P500 (Standard and Poor's 500) companies, the relative importance of the intangible assets over the total patrimonial value of those organizations has increased dramatically in the last 40 years, passing from representing about 20% (122B$ intangibles vs. 594B$ tangibles) in 1975 to a ratio of more than 5 to 1 in 2018 (21T$ intangibles vs. 4T$ tangibles) (Ponemon, 2019).

EU Valorization policy strongly encourages the use of knowledge and technology, the management of intellectual property, and the involvement of citizens, academia, and industry (EU valorisation policy, 2020; Demarinis et al., 2022).

Today, intellectual property assets are both engines of development and drivers of social transition. Industries that make intensive use of intellectual property rights (IPRs), such as patents, trademarks, industrial designs and copyrights, generate 45 percent of annual GDP (€6.6 trillion) in the EU and account for 63 million jobs (29 percent of all jobs) (EU valorisation policy, 2020).

In recent years, Intellectual Property Analytics (IPA) has emerged as a multidisciplinary approach used to gain valuable insight about intellectual property data (Trippe, 2015; Aristodemou & Tietze, 2018).

Of course, patent data can be analyzed in a variety of ways to fulfill different purposes. Organizations analyze patents for, but not limited to, determining novelty in technologies (Erzurumlu & Pachamanova, 2020; Bonino, Ciaramella & Corno, 2010), forecasting technological developments in a particular domain (Ernst, 2003) and technological road mapping (Phaal, Farrukh & Probert, 2004), analyzing patent trends, strategic technology planning, extracting the information from patents for identifying the infringements, determining patents quality analysis for R&D tasks, identifying the promising patents, the technological vacuums, hotspots, and technological competitors (Abbas, Zhang & Khan, 2014).

Besides, intellectual property knowledge databases come in heterogeneous forms (text, data, images, colors, and smells), and it is becoming increasingly difficult to analyze, synthesize and classify their contents (Trappey, Lupu & Stjepandic, 2020).

Often the classification of patents and, therefore, the search and consultation method, are based on taxonomies self-defined by experts or database managers and are not very effective.

Automated approaches, such as natural language processing for data, text and graph mining, clustering and neural networks, are increasingly used for IP knowledge processing and various tools have been developed for supporting patent analysis experts, business managers, and technology offices (Trappey et al., 2009; Lei, Qi & Zheng, 2019; Yoon & Park, 2007; Rodriguez et al., 2016; Puccetti, Chiarello & Fantoni, 2021; Kang et al., 2020; Trappey et al., 2020; Song, Ran & Yang, 2022).

Analyzing patent data using the automated tools to discover the patent intelligence through visualization, citation analysis, and other techniques, such as text mining is termed as "patent informatics" (Trippe, 2003). These techniques can be broadly classified into text mining techniques and visualization techniques and involve several steps, including extracting patents from databases, extracting information from patents, and analyzing the extracted information to derive logical conclusions (Abbas, Zhang & Khan, 2014).

Figure 1 shows a generic workflow that can be used for patent analysis.

In this scenario the aim of this paper is twofold:

- To define a specific patent analysis workflow aimed at improving patent classification and fruition;

- To experiment the proposed workflow on an Italian database named “Knowledge-Share” (KS).

As side effect the obtained results can also outlines the presence of different technology domains (or “technological areas”) with respect to those ones originally proposed by the KS platform and thus suggesting the use of a different taxonomy for patents classification.

The research questions addressed are:

- is the taxonomy based on 10 technological areas used by KS platform effective in classifying the whole landscape of academic patents collected?
- to which extent the new taxonomy obtained overlap with those one used by the KS platform?

The paper is organized as follows: Section 2 illustrates the materials and methods (Knowledgeshare database and the proposed approach with NLP and clustering), Section 3 describes the analyses and the results, Section 4 draws some conclusions and suggests future developments.

2 MATERIALS AND METHODS

2.1 The Knowledgeshare Database

Very often university patents remain unvalued and unexploited. They remain stuck in what is called the “Valley of death” which represents “the gap between where publicly available research funding stops and where private investment or commercial funding starts” (Hockaday, 2020). Recently, many universities tried to enhance the value of their research results to a greater extent, increase the valorization of research results and the market uptake of innovative solutions.

Improving easier access to and sharing of intellectual assets is “key to increase the valorization of research results and the market uptake of innovative solutions”: this is one of the objectives envisaged in the IP Action Plan 2020 of the European Commission (EU valorisation policy: making research results work for society, 2020). The ASTP Survey Report on KT Activities FY2019 shows a significant lack of valorization of the patented technologies from the Universities and Public Research Centers across Europe, since only 18% of those inventions are licensed or optioned (ASTP Report, 2019).

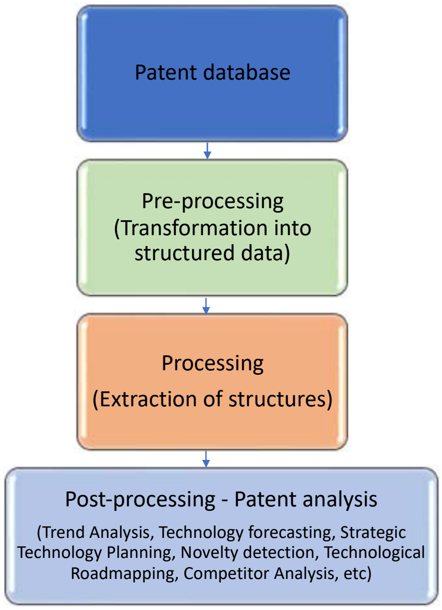

Figure 1: Workflow of patent analysis (from patent database to pre-processing, processing and post processing in patent analysis).

The Italian landscape, as a part of this picture, suffers from the same lack of exploitation. In order to overcome the difficulties that many Italian universities face in effectively promoting their research results and their IP assets, the Knowledge-Share platform was developed as a joint project involving Politecnico di Torino, the Italian Patent and Trademark Office at Ministry for Economic Development and Netval (the Italian Network for the Valorization of Public Research). KS is a platform designed for Italian Universities, Research Centers (PROs), Scientific Institute for Research, Hospitalization and Healthcare (IRCCS) to showcase their patented technologies and spin-off projects seeking commercialisation opportunities, and for businesses to find solutions and expertise to overcome R&D&I challenges (Technology Transfer System Handbook, 2019). It is specifically aimed to “translate” the contents of academic patented inventions into a self-speaking language which anybody can understand (the so called “patent marketing briefs”), thus obtaining three important results: i) to generate a real social and economic impact at national level, in accordance with the objectives of the “Third Mission”; ii) to provide a tangible support to (not only) Italian businesses to accelerate their innovation processes; iii) to drive economic return for Universities and PROs to be re-invested in new technology transfer activities within the public research system.

Particularly the platform’s key objectives are to:

• Become the touchpoint between corporations, SMEs and public research;

• Create a national standard to foster the exploitation of intellectual property;

• Create an innovation network for technological excellence at an international level;

• Provide industry scouting teams with an easy and effective way to tap into the Italian research landscape;

• Provide a service for technology transfer offices (market intelligence);

• Promote and foster events and initiatives related to innovation and exploitation of research;

• Generate spin-offs and innovative technology projects.

The KS database contains 1694 patents, uploaded on the platform by 89 Research Institutions (Universities, Research Centers, Scientific Institute for Research, Hospitalization and Healthcare, etc). The application date of the patents ranges from 1999 to 2021. The patents are categorized according to a taxonomy based on10 technological areas: Aerospace and aviation, Agrifood, Architecture and design, Chemistry, Physics, New materials and Workflows, Energy and Renewables, Environment and Constructions, Health and Biomedical, Informatics, Electronics and Communication System, Manufacturing and Packaging, Transports. The most populated technological area is Health and Biomedical that includes more than 1/3 of the total number of patents in KS database.

The website (https://www.knowledge-share.eu/, last accessed 16/12/22) allows the search for patents according to several criteria: the name, the organization it comes from (i.e. the patent owner), the area of application, the events at which they were shown and free full text research. For each patent, described in an informative and non-technical language, the technical features, the applications, uses and characteristics, and the benefits deriving from the adoption of the technology are illustrated. Furthermore, information about the inventors, the priority number, the priority date, the license, the commercial rights can be found and it is possible to download a “marketing annex”, i.e., a sheet that contains the basic information on the patent, conceived to be a functional and brief communication tool to share and circulate outside the platform (Fig. 2).

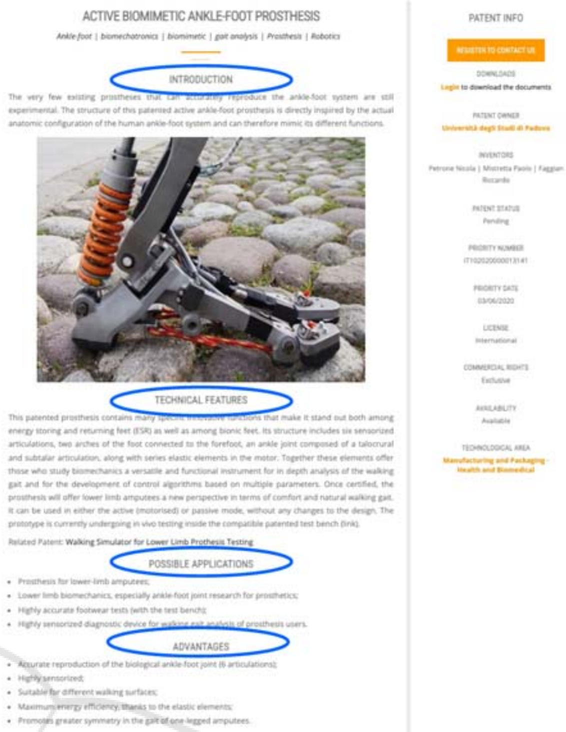

Figure 2: Marketing annex of each patent uploaded on Knowledge-Share platform.

There are currently more than 1520 registered users of the KS platform, consisting of companies, investors, banks, stakeholders and so on. In recent years, more than 250 contacts have initiated between universities and companies and some of them have already led to signed contracts and multiple forms of collaborations (co-development agreement, new research agreement leveraging on the same know-how of a particular invention, license or option agreements, etc.).

The main challenges now faced by the KS platform concern:

1. The improvement of the classification system to enhance the retrieval of relevant information.
2. The design of pre-matching strategies to provide an automated or semi-automated user support.
3. The identification of trends and forecasting of emerging technologies.

In this paper we propose a focused patent analysis workflow to face the first challenge.

2.2 Mining the KS Data Base

The Figure 3 shows the proposed approach for patent analysis and its application to KS. Natural Language Processing (NLP) techniques have been applied on the "marketing annex" of the 1694 patents, and in particular the sections "introduction", "technical features" and "application" were processed. These fields include all the relevant information about the patent without substantial redundancies.

We performed some of the main steps of NLP in the following order:

- Tokenization.
- Stop-words removal.
- Stemming.

In the *Tokenization* phase, texts were subdivided into single words (also called "tokens").

In the *Stop-words removal* phase all the useless token (such as articles, prepositions conjunctions, punctuation, numbers etc.) were removed.

In the *Stemming* phase, the remaining words were "stemmed" so that only the root-words were kept; for example, "fished" and "fishing" were transformed into their common root-word "fish".

Accordingly, we applied NLP techniques to our Italian corpus (i.e., the whole set of our collected texts).

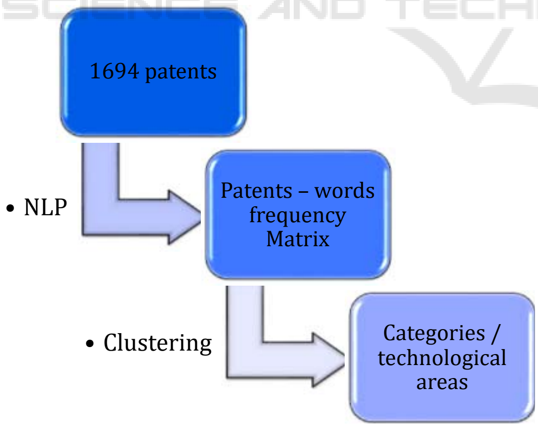

Figure 3: Proposed workflow for KS patent analysis.

Afterwards, we built the TF-IDF matrix (Beel et al., 2016). The rows of the TF-IDF matrix represent the texts of the corpus, while the columns stand for the words. The generic element of this matrix is a product of two terms:

$$(TF-IDF)_{ij} = (TF)_{ij} \times (IDF)_{ij};$$

where

$$(TF)_{ij} = \frac{f_{ij}}{\sum_k f_{ij}} \quad ;$$

$$(IDF)_{ij} = log \frac{N}{\left|\{d \in D | w_j \in d\}\right|} \quad ;$$

where:

- f_{ij} = frequency of word j in patent i;
- N = total number of patents in the corpus;
- D = set of patents, so that |D| = N;
- w_j = j – th word;
- d = document in D.

We obtained a matrix of 1.694 rows (patents) and 12.505 columns (words). It should be noted that, as pointed out in the literature (Jun, Park & Jang, 2014), this matrix is sparse.

Accordingly, in order to reduce the sparsity and make the clustering process less prone to the curse of dimensionality, we applied the Singular Value Decomposition (SVD) (Abdi, 2007) for dimension's reduction. Finally, we used the k-means clustering algorithm (Trappey, Trappey & Chung, 2017; Bock, 2007; Trappey et al., 2013).

Since both SVD and k-means depend on user-defined parameters, optimal values for these parameters must be chosen. In particular, since SVD shrinks the dimension of matrices by using linear combination of columns, it must be decided the optimal number of these linear combinations; for what concerns k-means, the most important parameter to choose is the number of clusters to retrieve. Accordingly, we performed a grid-search exploration of the parameters' space and used the Silhouette (Jun, Park & Jang, 2014) for optimization.

Following its definition, Silhouette is a combination of two terms: (1) the mean distance of each point of a cluster to the other points of the same cluster; (2) the mean distance of each point of a cluster to all the points in different clusters. Nonetheless, since there are many types of distance function that can be used in computing Silhouette, we considered three different distance measures (Euclidean, Manhattan and Cosine) and chose the function maximizing the Silhouette for any given value of the SVD and k-means parameters. Therefore, the grid-search explored three quantities: (1) the number of linear combinations in SVD; (2) the number of clusters in K-MEANS; (3) the distance function in Silhouette. The combination having the maximum value of Silhouette was the following:

- SVD - number of linear combinations 10,
- K-means - number of clusters 8,

- Silhouette - cosine distance.

Thus, the results obtained, presented in the next section, are based on this configuration of parameters.

3 RESULTS AND DISCUSSION

The first research question addressed here deal with the number of technological areas, i.e. the number of class in the taxonomy used by KS. Are these 10 areas sufficient to capture the whole landscape of academic patents? If not, do the data suggest a different classification?

Fig. 4 shows that according to Silhouette 8 clusters should be considered.

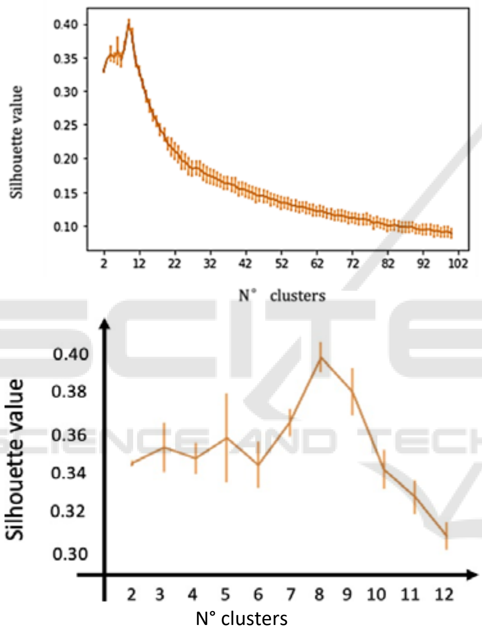

Figure 4: The variation in Silhouette value as a function of the number of clusters (the second one shows a magnification).

To fully answer the previous research question, we examined each cluster trying to evaluate whether there was a consistent overlap with the KS classification. Besides, we examined the keywords (in terms of frequency) for each cluster, and we assigned to each cluster a coherent although subjective classification, Table 1 shows the clustering results and Figure 5 illustrates the wordcloud for the 8 clusters.

Table 1: Clustering results.

	CLUSTER	N° patents
1	Technologies 4.0 (mechanics and robotics)	271
2	Material science	341
3	Cancer treatment	108
4	Optics - Image processing	179
5	Sensor technology - ICT	288
6	New molecules - new compounds - pharmacology	242
7	Energy/green Technologies	114
8	Biomedical	133

Therefore, the available 1694 patents were finally grouped into 8 clusters.

The first cluster, observing the most frequent words that are system, robot, device, material, biomedical, sensor, etc., seemed to be related to technologies 4.0 including mechanics and robotics applied to different field: healthcare, transport, design, manufacturing, etc. In particular, a careful analysis of the patents contained in this cluster, in conjunction with the analysis of the most frequent words, reveals interesting innovation spanning different fields of research:

Figure 5: Wordcloud for the 8 clusters.

- automotive/transportation with patents pertaining to rail safety devices, underwater drones, innovations for cycling and motorcycling,

- aerospace with patents inherent in anti-icing systems for aircraft, rotating aerodynamic elements, drones, intelligent structures for integrity testing etc.

- architecture and design with innovations in construction for energy upgrading, seismic isolation or monitoring complex structures and in the fashion industry,

- healthcare which is the main area of application of the innovations in this cluster. The patents present are truly varied, ranging from innovative wheelchairs for the disabled, underwater guidance for the blind, innovative orthotics, endomedicular prostheses, sensorized heart valve prostheses to devices for radiotherapy, pneumatic muscles, artificial bladders, and innovative brassieres and sheaths. But the bulk of innovations revolve around robotics leading the way: robotic exoskeletons, robotic platforms for laparoscopy, wearable robots, biomimetic robots, robotic surgical simulator, robotic limbs etc.

The second one was dedicated to material sciences with focus on agrifood, chemistry&physics, manufacturing and environment sectors. In particular, a close analysis of the patents contained in this cluster and the most frequent words, reveals interesting innovations spanning different fields of research. For example, we find patents that create new technology for the production of composite ceramic powders, biocompatible sandwich panel, invention employing supercritical carbon dioxide to pasteurize foods or the realization of a natural product from Rosa canina seeds obtained by CO_2 extraction in supercritical phase. Other innovations concern the synthesis of a pesticide nano-formulation from environmentally friendly materials or high mechanical performance materials from stone processing waste, micro-algal photo-bioreactor, inhibitor preparation of unpleasant odors from household waste, an economical and effective method for treating wastewater in liquid form or a multifunctional hybrid material based on natural clays for environmental recovery and bioremediation. Many patents deal with the food sector: intelligent, active and biocompatible label; process for manufacturing additives for use in making antibacterial nanocomposites; innovative coating composed of a glassy matrix and metal nanoparticles with antiviral, antibacterial and antifungal properties; ready-made base for chocolate confectionery products; device for removing proteins, metals and other instability agents from wine and vegetable beverages; production of biopolymers, exploiting agro-industrial wastes, etc.

The cluster n.4 was focalized on optics and image processing; in fact we can find here innovations about packaging of optical signals, automatic machine for real-time detection of contaminants, optoelectronic apparatus for measuring position and orientation of rigid bodies, x-ray concentrator, energy conversion device, solid-state photodetector, integrated optical device, automatic immersion ultrasonic system, new sensor for accurate pH measurements, multi-modal optical fiber communication system, fully automatic optical microscope for fast reading of samples consisting of a transparent dielectric with metal nanoparticles, use of porous silicon membranes decorated with silver or gold nanoparticles embedded in a microfluidic chip. In this cluster are also present some patents with medical application: textile electrode device for acquiring electrophysiological signals from the skin; video-assisted dentistry using intraoral video cameras; acquisition of surface electromyographic signals and ultrasound images from the same portion of muscle; measuring device for assessing the volume of a breast; treatment of tumors with ion beams (hadrontherapy).

Analyzing the most frequent words of the cluster n.5, we could deduce that it is concentrated on sensor technology and ICT in general. In fact, the main innovations are: soft-computing techniques for aerospace, anthropogenic noise control device; low-cost portable apparatus for characterizing sensor devices integrated with RFID transponders; RFID system developed for precise localization and tracking of objects equipped with low-cost tags; device prepared for analysis, simulation and prediction of slope instability phenomena; invention to protect files from ransomware-type attacks; ground-based synthetic aperture radar capable of acquiring both three-dimensional and two-dimensional images; virtual sensor organized with a trained neural network. There are also patents with healthcare application (such as wearable haptic system to guide the cadence of steps in a person through vibrotactile stimuli), agricultural application (technology for disinfection of agricultural soil through the use of microwaves or radio frequency), architectural application (innovative IoT system to manage building cooling through advanced machine learning techniques), automotive application (logic control system for automotive; platform, and related method, for the identification, classification and subsequent removal of manufacturing defects present on vehicle components; innovative system that projects vehicle information and augmented reality elements to enhance the driving experience) and economic ones (dynamic and responsive computer model capable of representing economic interactions among financial institutions).

Cluster n.7 is the one that best matches Knowledgeshare's classification; in fact, it is characterized by patents in energy and green technologies. Following are some examples: energy harvester device built into the stem of the paddles; energy converter, which uses gyroscopic effects to generate electricity from sea waves; energy conversion device, which allows electricity to be generated from wave motion; micro-wind generation system; enthalpy heat exchanger; solar-derived thermal energy storage and/or exchange device; motorized system with parallel kinematics that enables automatic cleaning of the surface of photovoltaic panels; spectrally selective solar absorber coatings with enhanced photo-thermal performance and stability; device that produces water from the air; innovative portable device capable of ionizing water taken from environments outside the home; new method for the simultaneous treatment of polluted water and power generation; system to carry out air humidification and heat recovery in air conditioning systems; electric generator that statically converts heat into electricity without the use of moving parts or matter flows; heat and mass exchanger made from biocomposite hydrogel. Other innovations go towards the automotive application: new cooling solution applicable to electric machines; integrated system capable of transforming an internal combustion vehicle into an electric vehicle; hybrid-electric light aircraft and so on.

Three clusters (n. 3, 6 and 8) were totally devoted to the health and biomedical sector, broadly understood.

The cluster n.3 was more related to cancer treatment with very relevant innovation in this field, such as: photodynamic therapy as a promising noninvasive treatment for cancer and nonmalignant tumors; method for early diagnosis and/or monitoring of Mucor infection; tumor suppressor of malignant mesothelioma of the pleura; innovative theranostic system involving a multifunctional nanoconstruct and ultrasonic activation set-up capable of treating cancer cells; use of particular strawberry extracts for the prevention, treatment, and/or control of the progression of uterine fibroids; circulating bio-marker for diagnosis and prognosis of tumor progression; multi-modular and innovative system capable of isolating stem cells from small amounts of adipose tissue; molecular markers predictive of response to immunotherapy; novel method for cryopreservation of dental pulp to isolate mesenchymal stem cells; efficient targeted delivery system of molecules with therapeutic action (e.g., cytotoxic agents) based on adipose stromal stem cells; method for identifying a biomarker of stemness in hepatocellular carcinoma cells. Some patents were focalized on regenerative medicine: non-erodible, sterilizable, biocompatible hydrogel scaffold for 3D cell cultures or recombinant protein scaffold for preparing cell culture plates for use in developing biomaterials for neuro-regenerative medicine.

The cluster n.6 was more connected with the formulation of new compounds and the discovery of new molecules and therapies. Some examples: medicine designed to counteract the progression of acute and chronic neurodegenerative diseases; RNA interference-mediated therapy for neurodegenerative diseases; synthesis of a leptin antagonist tetrapeptide; synthetic melanocortins with antimicrobial effects for the treatment of topical infectious diseases; pharmaceutical compound for the treatment of wounds for topical use; pharmaceutical composition, which contains bactericidal/permeability-increasing protein and hyaluronic acid with the purpose of treating different types of arthropathies; new peptide and its use for the treatment of Alzheimer's disease; nanostructures capable of delivering oxygen into hypoxic tissues, which are associated with various metabolic, ischemic, and infectious diseases; multifunctional biomaterial consisting of a hydrogel (hydrogel) that is administered through an injection directly into the tissue to be treated; use of irisin, a hormone secreted mainly from skeletal muscle and in smaller amounts from adipose tissue, as a drug/strategy for the preservation of the function and survival of pancreatic islets of Langerhans and, in particular, the beta-cells that are deputed to insulin production. Other innovations deal with: antiviral compounds that find application in the prevention and treatment of infections caused by coronaviruses; use of benzofurans as synthetic "natural-like" herbicides, characterized by high phytotoxic and herbicidal activity; micro and nanocapsules tannins useful for the preparation of controlled-release pharmaceutical, nutraceutical or cosmetic compositions; new yeast strain that can be used to combat fungal infections in fungi of agronomic and commercial interest; lentil extract with cholesterol-lowering and prebiotic activity, particularly useful in therapeutic applications and used as a nutraceutical; mixture of active ingredients from pomegranate seeds useful in the treatment and/or prevention of obesity and associated diseases, such as particularly insulin resistance and type 2 diabetes and hepatic steatosis.

The last cluster (n.8) was named “biomedical” as it included mainly methods and techniques for disease diagnosis and monitoring. Some of these are very interesting: new diagnostic marker for Paget's bone disease and associated bone tumors; new diagnostic test to identify the two most common mutations in chronic myeloid tumors; new test for early detection

Table 2: Comparison between KS technological areas and clusters emerged by the proposed approach.

Knowledgeshare areas vs Clusters	**Technologies 4.0**	**Material science**	**Cancer treatment**	**Optics - Image processing**	**Sensor technology ICT**	**New molecules, compounds, pharmacology**	**Energy/green Technologies**	**Biomedical**	**TOTAL**
Aerospace, aviation et al	18	9		15	19	2	3		66
Agrifood et al	5	72		10	18	15	5	8	133
Environment and Constructions et al	54	65	1	17	40	1	26	1	205
Architecture and design et al	22	4		1	14		6	1	48
Chemistry, Physics, New materials and Workflows et al	18	169	6	76	8	34	11	9	331
Energy and Renewables et al	3	7		5	12		60		87
Informatics, Electronics and Communication System et al	33	1	1	22	140	2	1	5	205
Manifacturing and Packaging et al	17	6		3	2		1		29
Health and Biomedical	94	14	100	29	26	188	3	112	566
Transports et al	9			2	12		1		24
TOTAL	273	347	108	180	291	242	117	136	1694

of colorectal cancer which assesses decreased expression of a protein; noninvasive method suitable for pancreatic cancer diagnosis at an early stage; fecal sample testing system able to diagnose major chronic inflammatory bowel diseases; innovative system for early diagnosis of acute renal failure; method of in vitro diagnosis of head and/or neck cancer in tissue and/or biological fluid samples; new reporter system that enables early detection of the occurrence of muscle atrophy; diagnostic for rapid and early differential diagnosis of ulcerative rectocolitis; next-generation sequencing techniques to detect specific molecular signatures of urinary miRNA, which can be used to distinguish bladder cancer cases from healthy controls; use of low field nuclear magnetic resonance for monitoring patients with cystic fibrosis; innovative method that allows separation of nucleated fetal cells from maternal peripheral blood at all gestational ages, etc.

As a second research question, we evaluated to which extent the new classification overlapped with the one used by the KS platform. This is strategic if we want to highlight the presence of emerging technologies/fields or cross-domain areas. To answer this question, we built a "contingency table" to compare KS classifications (rows) with those arising from our analyses (columns), see Table 2.

Our findings highlight a mismatch between technological areas defined in KS, as a result of the self-assignment performed by the inventors, and the clusters emerged by applying the proposed approach.

This finding suggests that the patent analysis workflow defined can lead to an alternative and probably more coherent classification, improving the offer/demand matching. We are evaluating with Netval partner the possibility to change the category/technological areas fixed in KS and/or create a recommender system able to help the inventors to correctly assign the patents.

As previously said, it is interesting to note that patents related to the health sector, originally located in only one technological area in KS, "Health and Biomedical" category, that account for a large portion (about 30 percent of the total), after the processing proposed in this paper can be rearranged into multiple and more specialized categories. In fact, in addition to the three clusters directly related to health, namely No. 2, No. 6 and No. 8, Cluster No. 1 - Technologies 4.0 - contains several patents connected to new technologies applied to health. As future work, we started to extract the most frequent words from the 4 aforementioned clusters dedicated to healthcare and a complex weighted network was built on them; the aim is to identify a preliminary "vocabulary" dedicated to Healthcare 4.0 from the huge mine of information contained in the Italian patent heritage. In fact, Healthcare 4.0 represents a great challenge in Industry 4.0. Integrating innovative technologies such as the Internet of Health Things (IoHT), cyberphysical CPS (medical CPS) systems, health cloud, health fog, big data analytics, machine learning, blockchain, and intelligent algorithms, it has the goal of providing better, value-added, and cost-effective health care services to patients and improving interactions between patients, stakeholders, infrastructure, and the value chain (Al-Jaroodi, Mohamed & Abukhousa, 2020). Furthermore Healthcare 4.0 focuses on the implementation of integrated healthcare platforms with progressively virtualized, distributed and real-time healthcare services for patients, professionals and formal and informal care givers. It proposes a shift from a practitioner-oriented hospital model to a distributed, patient-centered model (Thuemmler & Bai, 2017), as well as the integration, sharing and optimization of the use of health service resources, practitioners and systems management to improve operations and reduce costs (Al-Jaroodi, Mohamed & Abukhousa, 2020).Thanks to this research stream, it will be possible to identify promising and emerging technologies to become useful applications and profitable products in healthcare 4.0 (Wang & Lin, 2022).

4 CONCLUSION

In this work we propose a specific workflow for patent analysis together with its experimentation on the “Knowledge Share” platform, an Italian patent database, that includes 1694 patents from 89 Research Institutions (Universities, Research Centers, Scientific Institute for Research, Hospitalization and Healthcare, etc.), classified in 10 technological areas. Using NLP techniques and clustering analyses, the 1694 patents were re-arranged into 8 clusters, namely: Technologies 4.0, Material science, Cancer treatment, Optics - Image processing, Sensor technology – ICT, New molecules - new compounds – pharmacology, Energy/green Technologies, Biomedical. Our findings highlight a mismatch between the technological areas defined in KS and the clusters obtained by using proposed workflow. This mismatch is probably due to the self-assignment performed by the inventors when the patents were uploaded on the platform. An automated classification could be more coherent and could therefore lead to better performance in terms of offer/demand matching.

The potential benefits are countless, ranging from the possibility for companies and investors to take advantage of innovations produced by research institutions through an improved matchmaking system to the possibility of introducing novel technological areas focused on the actual innovation landscape, thus easing the development of cross-domain technologies and/or emerging technologies.

This study shows that Italian patents represent an extraordinary source of innovation that unfortunately is not yet fully "exploited" as all these inventions do not reach the market and the population. A great effort is still needed for a more intelligent management of intellectual assets that are the only factor that can foster innovation, creativity, knowledge sharing and improve the chances that knowledge reaches the market and brings faster benefits to society. This is especially true and strategic for the Healthcare sector that represents a main challenge in the Industry 4.0 for the final improvement of quality life and wellbeing of community and territory.

In conclusion:

- this research aims to emphasize the importance of technology transfer in the fourth industrial revolution, from the perspective of the quadruple helix model which describes university-industry-government-public interactions within a knowledge economy;
- patent analysis represents a key tool for organizations able to analyze, synthesize data and documents for insights, forecasts, technology trends, technology infringement, IP related management etc;
- companies and investors could take advantage of innovations produced by research institutions through a matchmaking system provided by Knowledge share platform;
- “Knowledge share” platform aims at helping innovators and researchers to make the most of their results and inventions thereby generating societal impact, as recommended by the EC Intellectual Property Action Plan;
- the potential application of a quantitative framework for innovation are countless, ranging from scientific policy to R&D strategies for firms, regions and even nations; it can be connected to socioeconomic data, to products and can be embedded in frameworks for industrial development.

REFERENCES

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3-13.

Abdi, H. (2007). Singular value decomposition (SVD) and generalized singular value decomposition. *Encyclopedia of measurement and statistics*, 907-912.

Al-Jaroodi, J., Mohamed, N., & Abukhousa, E. (2020). Health 4.0: on the way to realizing the healthcare of the future. *Ieee Access*, *8*, 211189-211210.

Aristodemou, L., & Tietze, F. (2018). The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, *55*, 37-51.

Aristodemou, L., Tietze, F., Athanassopoulou, N., & Minshall, T. (2017). Exploring the future of patent analytics: a technology roadmapping approach.

ASTP Report (2019) https://www.astp4kt.eu//assets/documents/Report%20-%20ASTP%20Survey%20on%20KT%20Activities%20FY2019.pdf (last accessed 16/12/2022)

Baglieri, D., & Cesaroni, F. (2013). Capturing the real value of patent analysis for R&D strategies. *Technology Analysis & Strategic Management*, *25*(8), 971-986.

Bartoloni, S., Calò, E., Marinelli, L., Pascucci, F., Dezi, L., Carayannis, E., ... & Gregori, G. L. (2022). Towards designing society 5.0 solutions: The new Quintuple Helix-Design Thinking approach to technology. *Technovation*, *113*, 102413.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Paper recommender systems: a literature survey. *International Journal on Digital Libraries*, *17*(4), 305-338.

Bock, H. H. (2007). Clustering methods: a history of k-means algorithms. *Selected contributions in data analysis and classification*, 161-172.

Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, *32*(1), 30-38.

Demarinis Loiotile, A., De Nicolò, F., Agrimi, A., Bellantuono, L., La Rocca, M., Monaco, A., ... & Bellotti, R. (2022). Best Practices in Knowledge Transfer: Insights from Top Universities. *Sustainability*, *14*(22), 15427.

Ernst, H. (2003). Patent information for strategic technology management. *World patent information*, *25*(3), 233-242.

Erzurumlu, S. S., & Pachamanova, D. (2020). Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. *Technological Forecasting and Social Change*, *156*, 120041.

EU valorisation policy (2020) https://research-and-innovation.ec.europa.eu/system/files/2020-03/ec_rtd_valorisation_factsheet.pdf (last accessed 16/12/2022)

EU valorisation policy: making research results work for society (2020). https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/eu-valorisation-policy_en (last accessed 16/12/2022)

Hockaday, T. (2020). University Technology Transfer: What it is and how to Do it. JHU Press.

Jun, S., Park, S. S., & Jang, D. S. (2014). Document clustering method using dimension reduction and support vector clustering to overcome sparseness. *Expert Systems with Applications*, *41*(7), 3204-3212.

Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139-172.

Lei, L., Qi, J., & Zheng, K. (2019). Patent analytics based on feature vector space model: A case of IoT. *Ieee Access*, *7*, 45705-45715.

Phaal, R., Farrukh, C. J., & Probert, D. R. (2004). Technology roadmapping—A planning framework for evolution and revolution. *Technological forecasting and social change*, *71*(1-2), 5-26.

Ponemon Report (2019) https://www.aon.com/getmedia/60fbb49a-c7a5-4027-ba98-0553b29dc89f/Ponemon-Report-V24.aspx (last accessed 16/12/2022)

Press release European Commission (2020). "Commission adopts Action Plan on Intellectual Property to strengthen EU's economic resilience and recovery", https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2187 (last accessed 16/12/2022)

Puccetti, G., Chiarello, F., & Fantoni, G. (2021). A simple and fast method for Named Entity context extraction from patents. *Expert Systems with Applications*, *184*, 115570.

Rodriguez, A., Tosyali, A., Kim, B., Choi, J., Lee, J. M., Coh, B. Y., & Jeong, M. K. (2016). Patent clustering and outlier ranking methodologies for attributed patent citation networks for technology opportunity discovery. *IEEE Transactions on Engineering Management*, *63*(4), 426-437.

Schwab, K. (2017). The fourth industrial revolution. Currency.

Song, K., Ran, C., & Yang, L. (2022). A digital analysis system of patents integrating natural language processing and machine learning. *Technology Analysis & Strategic Management*, 1-17.

Technology Transfer System Handbook https://www.polito.it/imprese/trasferimento/TTS_handbook.pdf (last accessed 15/12/2021)

Thuemmler, C., & Bai, C. (2017). Health 4.0: application of industry 4.0 design principles in future asthma management. In Health 4.0: How virtualization and big data are revolutionizing healthcare (pp. 23-37). Springer, Cham.

Trappey, A. J., Lupu, M., & Stjepandic, J. (2020). Embrace artificial intelligence technologies for advanced analytics and management of intellectual properties. *World Patent Information*, *61*, 101970.

Trappey, A. J., Trappey, C. V., & Chung, C. L. (2017). IP portfolios and evolution of biomedical additive manufacturing applications. *Scientometrics*, *111*(1), 139-157.

Trappey, A. J., Trappey, C. V., Chiang, T. A., & Huang, Y. H. (2013). Ontology-based neural network for patent knowledge management in design collaboration.

International Journal of Production Research, *51*(7), 1992-2005.

Trappey, A. J., Trappey, C. V., Hsu, F. C., & Hsiao, D. W. (2009). A fuzzy ontological knowledge document clustering methodology. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(3), 806-814.

Trappey, A. J., Trappey, C. V., Wu, J. L., & Wang, J. W. (2020). Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Advanced Engineering Informatics*, *43*, 101027.

Trippe, A. (2015). Guidelines for preparing patent landscape reports. *Patent landscape reports. Geneva: WIPO*, 2015.

Trippe, A. J. (2003). Patinformatics: Tasks to tools. *World Patent Information*, *25*(3), 211-221.

Wang, Y. H., & Lin, G. Y. (2022). Exploring AI-healthcare innovation: natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes*.

Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE transactions on engineering management*, 54(3), 588-599.