Using LDA for Innovation Topic of Technology : Quantum Dots Patent Analysis

Nurmitra Sari Purba[1], Rani Nooraeni[2]
15.8804@stis.ac.id[1], raninoor@stis.ac.id[2]

STIS, Politechnic of Statistics,,East Jakarta, 13330, Indonesia[1,2]

Abstract. This study seeks to explore information about one of nanotechnology, quantum dots (QDs), through analysis of patent information. QDs patent documents obtained from the United States international patent database, the USPTO, use web scraping. In total, 3914 patents from 1988 to 2016 were taken and archived for analysis. This paper discusses how to apply Latent Dirichlet Allocation (LDA), a topic model, in a trend analysis methodology that exploits patent information. After the text preprocessing and transformation, the number of topics is decided using the log likelihood value. Then LDA model is used for identifying underlying topic structures based on latent relationships of technological words extracted. We extracted words from 6 relevant topics and showed that these topics are highly meaningful in explaining technology applications of QDs.

Keywords: LDA, noun phrases extraction, patent map, quantum dots, text mining

1 Introduction

Mastery of technology is the key to the success of a nation. This can be witnessed from the success of Japan, South Korea and Taiwan which grew from developing countries to developed countries, even though they did not have sufficient natural resources. These countries have increased R & D activities to obtain the best and competitive technology.

The industrial revolution has evolved over 16 centuries, and as a result, the world economy has changed dramatically. Now, the world has begun to welcome the fifth era industrial revolution with its headline, "The fifth industrial revolution of nanotechnology". Even more striking is that nanotechnology is said to be able to reach the target of a country's economic achievement for a quarter of a century in just ten years (Rai and Rai, 2015).

Nanotechnology, a field that is prioritized and promoted by governments in almost all the world, is one of the fastest growing research fields in the scientific and technical fields (NSTC, 2006). With nanotechnology, materials can be arranged in order of atoms per atom or molecule per molecule so that no waste is not needed. By rearranging or engineering the material structure at the nanometer level, the material will have far more special properties that outperform the previous material. This is the background of countries in the world to compete to allocate funds for the development of nanotechnology. (BPPI, 2008).

One form of nanotechnology is quantum dots. By definition, QDs (quantum dots) are nano meter (nm) semiconductor materials, usually between 3 and 25 nm. When compared with the distance between atoms in a crystal arrangement, which is 1-2 angstroms (or 0.1 - 0.2 nm), the quantum dot (QD) consists of only about 1000 atoms. According to existing

ICSA 2019, August 02-03, Bogor, Indonesia

DOI 10.4108/eai.2-8-2019.2290336

literature, the electronic characteristics of QDs are determined by their size and shape, which means we can control their emission wavelengths by tuning their size. Their highly tunable optical properties based on their size are fascinating, leading to a variety of research and commercial applications including bioimaging, solar cells, LEDs, diode lasers, and transistors. The properties of quantum dots have led researchers to engage in various innovations. Increasing demand from various sectors has led to the growth of the quantum dots market which offers opportunities worth $ 2.76 billion in 2018, and the market value will balloon with a compound annual growth rate (CAGR) of 27.34% during the forecast period 2019-2025.

Patents are one of the Intellectual Property (IP) which contains comprehensive information about technological developments. Patent documents are important sources of information in science, technology, business, economics, law, etc. According to statistics from the World Intellectual Property Organization, 90-95% of world inventions can only be found in patent documents and 80% of these techniques do not appear in other professional articles (Liu and Shyu 1997). Patent analysis is a good tool for planning Research and Development (R & D), knowing new emerging technologies, analyzing technology trends, predicting technological developments, technology road maps, identifying technologies that are vacuum and which are in demand in industry and research, and identify technology competitors (Assad, 2013). Many researchers believe that knowledge about the development and diffusion of technology can be obtained through analysis of patent information, because patent documents are an important source of information for analyzing innovation in the field of technology if the data is analyzed systematically (Purba and Nooraeni, 2019).

Technological topic segmentation/recognition is one of the most important tasks for patent competitive intelligence analysis, since detecting technology trends, hot spots, and core technology are all based on it. Technological topic segmentation is dividing one research field into many sub-fields. Each sub-field consists of multiple technical words. In bibliometrics analysis, keyword occurrence, co-occurrence and co-citation analysis are often adopted for identifying subjects (Lee 2008; Hofer et al. 2010; An and Wu 2011; Zhang et al. 2012). As for many patent databases do not provide citation information, people often focus on co-word analysis for patent subjects identification. Keywords are highly generalized and concise words for representing papers, through keyword cooccurrence analysis one could detect research themes, clusters and knowledge structures etc. But for patent literatures, there are no keywords provided by patent applicants.

In this paper, topic modeling with the LDA method is used to identify latent topics hidden in patent documents. The topic model of LDA is a text mining method that is useful for detecting latent topics from large datasets. Although there are many studies related to text mining in patent documents (Rahmawati et al. 2017; Feng and Fuhai, 2012; Yoon et al. 2011), only a few of these studies apply the topic modeling method. There are several limitations in using LDA to find topics in patent documents. First, the number of topics is not automatically obtained in the LDA process. The researcher must set a number of topics right before running the LDA. Second, we must extract keywords that truly represent the discussion of patent documents. Considering the unavailability of keyword information in patent documents, this paper will use noun phrases taken from the title and abstract.

2 Materials

Quantum dots patent data will be taken from the USPTO website. Among national patents, USPTO patents are considered the most valuable because of the competitiveness of the US market. The US is a world leader in most technologies (Criscuolo, 2006). As a technology indicator, US patents can be considered the most reliable because companies want to secure their intellectual property rights in this largest market. From WIPO it is known that almost all patent applicants from various countries submitted their applications to the United States of America.

Patent documents can be accessed online on the USPTO website. However, the data is not available can be downloaded directly because it is in the form of web pages. Therefore, a web scraping technique for data collection is needed that can extract patent documents from the USPTO website. Web scraping extracts the quantum dots patent document with the specified search query:

(ACLM / "quantum dots" OR ACLM / "quantum dot" OR ACLM / "quantum-dot" OR ACLM / "quantum-dots") and (APD / 1 / 1 / 1976-> 12/31/2016)

3Methods

3.1 Patent map analysis

Patents filed in most patent databases contain various information such as publication dates and applications, applicants, inventors, and international classification numbers. Analysis of patent map information utilizes that information to make general summaries. One such summary is the number of patents that can be expressed as the number of cumulative patents or as the number of annual patents. The cumulative patent count reflects the technology life cycle which in turn can be used to determine the stage of technological development. If the analyst knows the stage of technology development, it is possible to estimate future trends and predict market saturation. Knowledge of the maturity and future growth of the market from technological innovation helps researchers decide whether to continue to invest resources or move towards research.

Patent map is an important problem in the analysis of patent information. The most popular map patents are built by extracting bibliographic patent documents based on descriptive and visual statistics (Camus and Brancaleon, 2003; Fattori et al., 2003; Morris et al., 2002). The presentation of patent maps is widely visualized in visual form (eg tables, graphs, and charts).

3.2Text Preprocessing

Patent titles and abstracts will be used for topic modeling. Titles and abstracts are chosen because they are general summaries of patents containing patent details and specifications. The topic modeling process cannot accept text document input because the structure is irregular. Therefore, preprocessing and transformation stages are needed, which in turn the text is represented as a number in a structured data format so that it can be accepted as an input to topic modeling.

Preprocessing Title and Absract

Text preprocessing used for extracting interesting and non-trivial and knowledge from unstructured text data. Preprocessing in the title and abstract of a patent document consists of several steps whose order can be seen at the following flowchart.

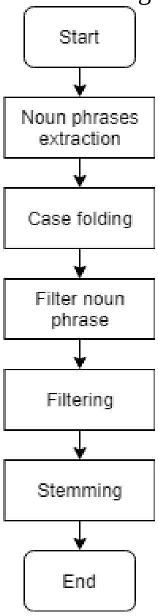

Fig. 1. Preprocessing flowchart

After the title and abstract are combined into one, the first step is extracting phrases and nouns. Compared to paper, the patent literature does not have keyword information. Defining a keyword series is very dependent on the efforts of experts, which may be expensive or not available (Yoon et al. 2011). Furthermore, determining keywords from technological innovations that contain new terms may also be a difficult task for experts.

According to the construction of sentences in linguistics, the main information of a sentence comes from nouns and noun phrases in a sentence. Therefore, this study will use the noun phrase extraction method to retrieve the keywords in the study. For extraction of noun phrases, in this case, it is necessary to make several rules for extracting noun phrases from patent documents. First of all, we must label part-of-speech for all the words in the patent document. Second, make rules based on part-of-speech from each word. Then the noun phrases will be extracted from the title and abstract of the patent document.

The following are examples of sentences with part-of speech.

a) Alignment systems for optical fibers attached to optical waveguides

b) Alignment / NNP system / NN for / IN optical / JJ fiber / NNS attached / VBN to / TO optical / JJ waveguides / NNS

The first sentence (a) is the original sentence, and the second sentence is the output of the Stanford Log-linear Part-Of-Speech Tagger. The letter after the ‘‘ / ’tag represents the group of words. For example, NN represents a noun, JJ represents an adjective. Based on this, it is necessary to define the rules for extracting noun phrases based on the structure and features of phrases. Extraction rules are as follows:

1. length of noun phrases no more than 5 words.

2. construction form of noun phrases as follows.
Noun
Noun + noun + ...
Adjective + Adjective + ... + noun + noun + ...
3. match part-of-speech titles and patent abstracts based on the form of the tagging above. Matching noun phrases will be extracted as descriptive terms.

After extracting noun and noun phrases, the next step is case folding, which converts all letters to lowercase letters. After that, the next step is filtering nouns. This stage is needed because in the noun phrases extraction step there are no noun phrases that do not meet the criteria of descriptive terms. These words include: 'use thereof', 'concrete realization', 'above-described advantages', 'SEQ ID No', 'similar manner', and 'more charges'. By inspecting one by one on all noun and noun phrases, the researcher decides to delete noun phrases if they contain the words 'use', 'realization', 'advantages', 'seq', 'manner', 'charges' on one words that make up noun phrases.

After filtering nouns, filtering is done again to delete words in the form of non-alphanumeric characters such as punctuation and or irrelevant symbols, numbers, words that are less than three letters, and stopwords. Stopwords are non-descriptive words that can be discarded. In removing stopwords, the researcher uses the stopword list from the python package. However, because there are additional stopwords in the title and abstract of this patent, adjustments are made by creating an additional stopword list.

The last step in preprocessing is stemming, the stage of finding the root words. The stemming stage in this study uses the Porter2's Stemmer algorithm or also called Snowball Stemmer for English texts.

Transformation of Title and Abstract

Because the topic modeling process requires data in a structured form, it is necessary to transform the text into numbers by weighting the word. Every word that appears in the document will be weighted depending on the frequency of its appearance in each document. Word weight calculation using TF will produce a vector space model or called document-term-matrix. The form of the document-term-matrix can be seen in the picture below. The matrix element contains the frequency of the i term in the nth document.

Table 1. Document-term-matrix

	term 1	term 2	term 3	term 4	term 5	...	...	...	term i
doc 1	$f_{1,1}$	$f_{2,1}$	$f_{3,1}$	$f_{4,1}$	$f_{5,1}$				$f_{i,1}$
doc 2	$f_{1,2}$	$f_{2,2}$	$f_{3,2}$	$f_{4,2}$	$f_{5,2}$				$f_{i,2}$
doc 3	$f_{1,3}$	$f_{2,3}$	$f_{3,3}$	$f_{4,3}$	$f_{5,3}$				$f_{i,3}$
doc 4	$f_{1,4}$	$f_{2,4}$	$f_{3,4}$	$f_{4,4}$	$f_{5,4}$				$f_{i,4}$
doc 5	$f_{1,5}$	$f_{2,5}$	$f_{3,5}$	$f_{4,5}$	$f_{5,5}$				$f_{i,5}$
...									
doc n	$f_{1,n}$	$f_{2,n}$	$f_{3,n}$	$f_{4,n}$	$f_{5,n}$				$f_{i,n}$

3.3Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is the most popular topic modeling method today. LDA can be used to summarize, cluster, link and process very large data. The LDA will produce a weighted list of topics for each document (Campbell, 2014). The distribution used to get the distribution of topics per document is called Dirichlet distribution, then in the generative process for LDA, the results from Dirichlet are used to allocate words to documents for different topics. In LDA, documents are observable objects, while topics, topic distribution per document, classification of each word on a topic per document is a latent (hidden) structure, which is why this method is called Latent Dirichlet Allocation (LDA) (Blei, 2012) . According to Blei (2003), LDA is a generative probabilistic model of a collection of documents called the corpus. The basic idea proposed by the LDA method is that each document is represented as a random mixture on a hidden topic, in which each topic has characteristics determined based on the distribution of the words contained in it.

Topic modeling with Latent Dirichlet allocation (LDA) was carried out on the title and abstract of patent documents that had gone through the preprocessing and transformation stages. LDA uses document-term-matrix input to inform research and research topics in Quantum dots patent documents. The results of the LDA will represent the most appropriate distribution of topics for each QDs patent document, these results can then be used to see which topics are the main concern of researchers in the world.

There are several limitations in using LDA to find topics that are hidden in patent document. First, the number of topics cannot be given automatically in the LDA topic model. The researcher must specify the number of topics right before running the LDA. Second, you must extract keywords from the patent document for the LDA topic model. The second weakness has been overcome by the extraction of nouns from patent titles and abstracts so that the document now contains a relatively descriptive word for the patent. One method that can be used to find the optimal number of LDA topics is by looking at the plot between the number of topics and log likelihood. The optimal number of topics is when the topic's log likelihood value is greatest. The first weakness will be overcome in this way, by looking at the LDA log likelihood plot when the number of topics is 5, 6, 7, 8, 9, and 10.

After the LDA model is obtained, the next step is the interpretation of the results. Results interpretation is a stage that is carried out to analyze the proportion of topics in the corpus.

4Result and Discussion

Patent data collection is done by web scraping on English-language patent documents in the United States patent database regarding quantum dots. From 1976 to 2015, there were 3914 patent QDs with the oldest initial patent year of 1988.

4.1 Patent map

The development of the number of QDs patents in the world In Figure 23, it can be seen that the growth of the quantum dots patent has a positive trend until 2015, then declined dramatically in 2016. Since the quantum dots partama patent was published in 1988 by the USPTO, quantum research and research has experienced very slow output growth over the next 20 years. Until finally in 2001, there was a significant surge and became the beginning of the increasing number of findings produced in this field. Growth in 2001 alone amounted to 49.16% and the increase was the biggest increase throughout the year of observation in this study. The decline in the number of patents in 2016 signaled that new innovation activities for quntum dots began to decrease.

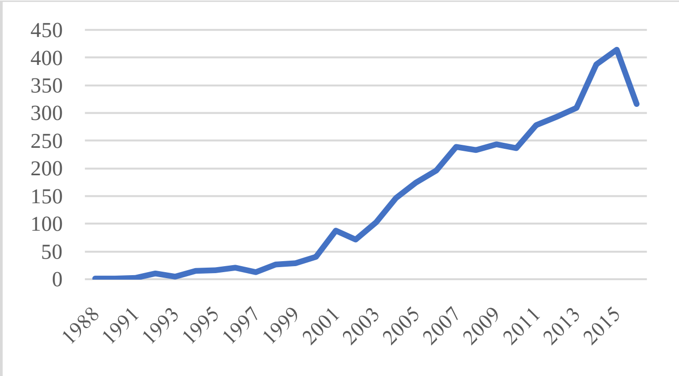

Fig. 2. Number of QDs patent grants per year in USPTO database

The development of the number of patent QDs per year in five main countries If observed from the five largest patent countries, namely the United States, South Korea, Japan, China and Taiwan, the same trend conditions were also experienced, where the quantum dots patent application declined simultaneously in the year 2016. From graph 4.5, it can be seen that the source of the big surge in 2001 came from research and research activities in the United States. 2001 was the starting point for the United States to grow and leave other countries in research in the field of quantum dots technology.

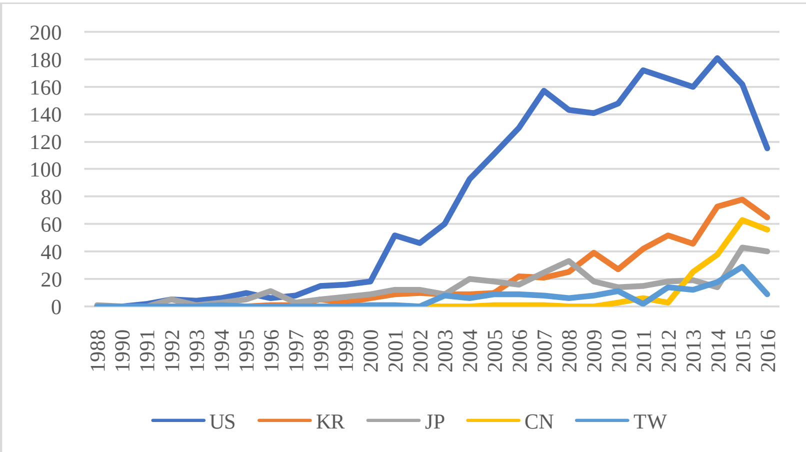

Fig. 3. Number of QDs patent grants per year in top 5 countries

Through the analysis of patent data from 1988 to 2016, information was obtained that there were 39 countries that applied patents from a total of 193 countries in the world. This gives the message that only a few countries are brave and willing to take the opportunity and opportunity to implement quantum dot in the industry and in research. The United States (US), with a percentage of 41.6% or as many as 2127 patents of all patents in the world, is the country that leads the mastery of quantum dots patents. In fact, aggregates from major players from Asia (South Korea (KR), Japan (JP), China (CN), Taiwan (TW)) were unable to keep up with the number of patents from the United States. It can be said, the United States is the most influential and dominant country in quantum dots technology. When viewed from many patents, the United States has no threat or a strong competitor from other countries. In Europe, Germany, Britain, the Netherlands and France play a role in patent quantum dots but they lag far behind the Asian and American continental countries. Another thing with countries on the African continent, no African country holds a quantum dots patent. These conditions can be seen clearly on the following map.

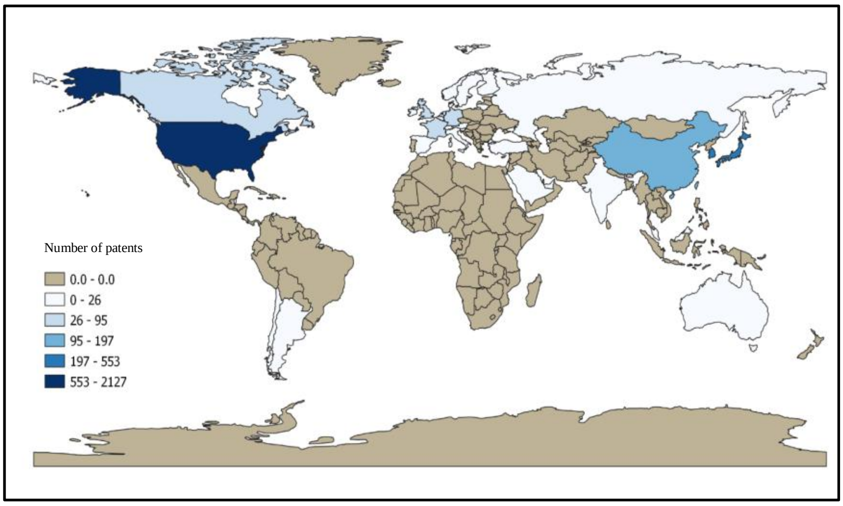

Fig. 4. Distribution map of QDs patent assignee

First of all, the text preprocessing process is first applied which will filter out unnecessary text data as explained in the methodology chapter. This stage succeeded in reducing the number of words from the initial 427550 to a very drastic decrease to 79041, or if it was suspended a number of 81.51% of the initial number of words.

Next, document-term-matrix (DTM) is formed from titles and abstracts that have gone through the preprocessing stage. Each row and column element is the frequency of occurrence of the word i in the mth document. DTM has a fairly high dimension because every single string of words counts as one dimension (3914 rows x 1718 columns).

Table 2. Document-term-matrix of cleaning text

	abil	absenc	absorb	⋯	nitrid	nitrogen	nois	⋯	zinc
doc1	1	0	0		0	0	0		0
doc2	0	0	0		0	0	0		3
doc3	0	0	0		0	0	0		0
doc4	0	0	3		0	0	0		0
doc5	0	0	0		0	0	0		0
⋮				⋱				⋱	
doc3914	0	0	0		0	0	0		0

As an unsupervised machine learning method, LDA requires defining the number of topics to be detected. The fewer number of topics indicate that the topic is too general and intermingles with one another. The more the number of topics, the topic will be too specific so that it contains words that cannot be concluded. To optimize LDA performance in finding topics, the likelihood log score is used to find the best number of topics. The topic selection range in this study is between 5 and 10 topics. For the title and abstract of the QDs patent, the log likelihood score shows that the six topics are the best number of topics based on the largest log likelihood value as shown in the figure below.

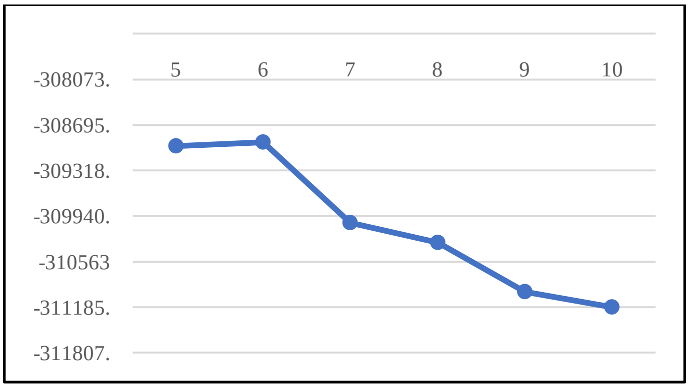

Fig. 5. Plot of log likelihood value per topic

Modeling topics with LDA is done with six topics. The six topics are then visualized with the PyLDAvis package in the python programming language. PyLDAvis offers the best visualization to see the distribution of keyword topics. A good topic model will have lumps that don't overlap, large enough for each topic. This happened in the distribution of topics in this modeling. It can be concluded that the model formed is good. The visualization of the results from topic modeling using the LDA is shown in Figure 31.

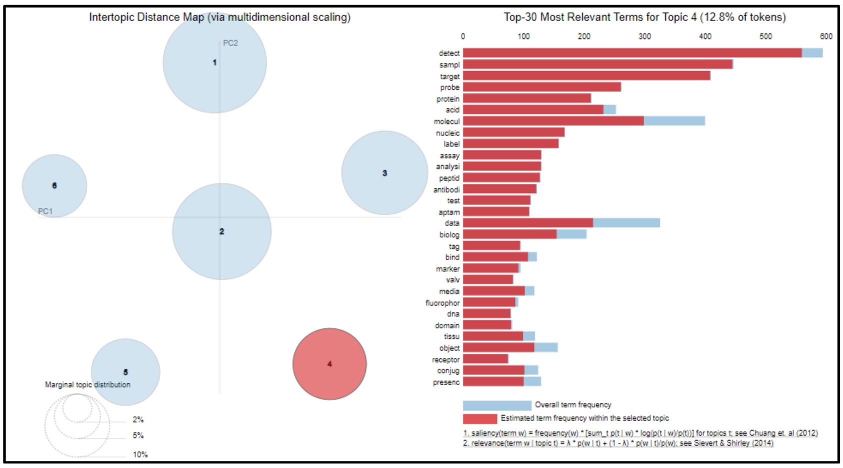

Fig. 6. Visualize the results of topic modeling with PyLDAvis

The circle in the left panel will refer to the topic, and the bar chart in the right panel will show the 30 most relevant words for the selected topic. Red bars represent the frequency of words in a particular topic, and blue bars represent the frequency of words in the entire corpus.Word probability distribution in six Topics (10 words per topic) from the model can be seen in the table below.

Table 3. Top 10 keywords on each topic

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
light	optic	composit	particl	layer	detect
display	laser	core	cell	substrat	sampl
wavelength	photon	polym	crystal	first	target
sourc	energi	nanoparticl	liquid	semiconductor	cell
color	state	group	fluoresc	second	molecul
first	wavelength	shell	solar	electrod	probe
convers	signal	nanocryst	graphen	region	acid
second	waveguid	luminesc	emiss	film	data
emit	radiat	metal	carbon	electron	apparatus
optic	semiconductor	semiconductor	size	conduct	protein

The following table presents the probability values of each topic in a document. For example, the probability value of topic 1 in the first document is 0, the probability value of topic 2 in the first document is 0.84 and so on according to the number of topics.

Table 4. Probability distribution of documents per topic

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
Doc1	0	0,84	0,15	0	0	0
Doc2	0,01	0,01	0,01	0,73	0,25	0,01
Doc3	0,86	0	0	0,13	0	0
Doc4	0	0,79	0,1	0,1	0	0
Doc5	0	0,98	0	0	0	0
⋮						
Doc 3914	0,24	0,74	0,01	0,01	0,01	0,01

Next, document classification is done based on the maximum value of each topic probability in a document. Topic labeling is done by looking at words that have the strongest probability of representing the topic, which can be seen in table 3.

Table 5. Classification of quantum dots patent documents

Topic	Topic label	Number of document	Percentage
1	Display Device	810	20,7%
2	Laser	537	13,7%
3	Core/Shell Nanocrystal	391	10,0%
4	Liquid Crystals with Graphene	239	6,1%
5	Semiconductor device	723	18,5%
6	Biological application	530	13,5%
	Even distribution	684	17,5%

The first topic is display device. Quantum dots are now in great demand for TV sets or computer screens. QD improves the performance of light-emitting diodes (LEDs), which lead to a new design "Quantum Dot Light Emitting Diode". QD is very useful for display devices given its unique optical properties. QD presents colors that look more accurate and extraordinary. QLED is more power efficient, thinner, and more flexible.

The second topic is Laser. Quantum dot lasers are semiconductor lasers that use quantum dot as active laser media in their light transmitting region. Lasers made from quantum dots show device performance that is closer to a gas laser, and avoid some negative aspects of the performance of devices associated with traditional semiconductor lasers. Quantum dots laser has begun to be applied in medicine (scalpel laser, optical coherence tomography), display technology (projection, TV laser), spectroscopy, and telecommunications.

The third topic is the Nanocrystal Core / Shell. This topic focuses on making polymers and making nanoparticles.

The fourth topic is Liquid Crystals with Graphene. Graphene is a material which is followed by carbon atoms that form the lattice structure of a honeycomb-like atom (Efelina, 2015). As a completely new material, graphene has extraordinary properties, including high electron mobility reaching 200,000 cm2 / Vs, high electrical conductivity, thermal conductivity high, good optical transparency, and strength 200 times harder than steel and 20 times harder than diamond. One sheet of graphene with an area of 1 m2 weighs only 0.77 mg (Huss and All, 2010). Based on the advantages of the properties possessed by Graphene above, Graphene material can be utilized in various fields, including the fields of electronics, photonic fields, and in other fields.

The fifth topic is the semiconductor device. Semiconductor devices that use quantum dots on patent QDs such as memory devices and solar cells. QD makes high-speed memory devices with a large and durable capacity. The solar cell is a device that is able to convert light energy into electrical energy. Solar cells on the market are usually based on silicon. However, the average efficiency achieved by solar cells is only around 15%. QD solar cells can increase efficiency by up to 30%, and can last up to 10000 hours of use (Abdullah, 2009).

The sixth topic in quantum dots technology is the biological application. Semiconductor nanoparticles, another name for quantum dots (QDs), have attracted attention and are rapidly expanding into the field of biology to the medical field. QD based molecular sensing has been developed so as to achieve high sensitivity, high throughput, and multiplexing capabilities. With its many advantages, quantum dots can detect cancer cells properly and are also used to deliver drugs precisely to the destination point in the human body (Jamieson, et al. 2007).

5Conclusion

The growth trend of quantum dots technology is positive until 2015. However, in 2016 there was a decrease in the number of patents that indicates that the output of innovation has begun to decrease. From the data it is known that until 2016, there were 39 countries that held quantum dots patents from 193 countries in the world. The United States is the country with the most patents (41.6% of the total patents in all countries).

Modeling topics with LDA produces six topics on quantum dots technology. The six topics are display devices, lasers, core/shell nanocrystal, semiconductor devices, liquid crystals with graphene, and biological applications where display devices are the most dominant quantum dots technology topics.

Industry in Indonesia is expected to have a depth of view in order to start relying on high technology. So far, Indonesia tends to use low to medium technology. Utilization of high technology will provide added value and will invite research for its development. The government is also expected to directly appeal to industries in Indonesia to apply nanotechnology in the production process.

References

[1]Abdullah, M. Pengantar Nanosains. Bandung : Penerbit ITB (2009).
[2]An, X. Y., & Wu, Q. Q. Co-word analysis of the trends in stem cells field based on subject heading weighting. Scientometrics, 88(1), 133–144 (2011).
[3]Assad, L. Zhang, S.U. Khan. A literature review on the state-of-the-art in patent analysis. in World Patent Information(2014) (Elsevier), p. 1-11 (2013).
[4]Biological applications of quantum dots. Biomaterials 28 4717–4732 (2007).
[5]BPPI [Departemen Perindustrian, Badan Penelitian dan Pengembangan Industri]. "Road Map Pengembangan Teknologi Industri Berbasis Nanoteknologi". (2008).
[6]Camus, C. and Brancaleon, R. Intellectual assets management: from patents to knowledge. World Patent Information, Vol. 25 No. 2, pp. 155-9 (2003).
[7]Chen, Y.-H., Chen, C.-Y., & Lee, S.-C. Technology forecasting and patent strategy of hydrogen energy and fuel cell technologies. International Journal of Hydrogen Energy, 36(12), 6957–6969 (2011).
[8]Criscuolo, P. The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and EPO. Scientometrics, 66(1), 23-41 (2006).
[9]D. M. Blei. Latent dirichlet allocation. Machine Learning Research 3, pp. 933-1022 (2003).
[10]D. M. Blei. Probabilistic Topic Model. communications of the acm, vol. 55 (2012).
[11]Efelina, V. Pengaruh Konsentrasi Urea dalam Sifat Optik Nanofiber Graphene Oxide/PVA (polyvinyl Alcohol) yang difabrikasi Menggunakan Teknik Electrospining.Tesis.Yogyakarta : FMIPA UGM (2015).
[12]Fattori, M., Pedrazzi, G. and Turra, R. Text mining applied to patent mapping: a practical business case. World Patent Information, Vol. 25, pp. 335-42 (2003).
[13]Feng, Xu & Fuhai, Leng. Patent text mining and informetric-based patent technology morphological analysis: An empirical study. Technology Analysis and Strategic Management 24(5):467-479 (2012).
[14]Hofer, K. M., Smejkal, A. E., Bilgin, F. Z., & Wuehrer, G. A. Conference proceedings as a matter of bibliometric studies: The Academy of International Business 2006–2008. Scientometrics, 84(3), 845–862 (2010).
[15]Huss, E. & All, F. Graphene. The Royal Swedish Academy of Science (2010).
[16]J. C. Campbell, A. Hindle and a. E. Stroulia. Latent dirichlet allocation: Extracting Topics. (2014).
[17]Jamiesona, T., Bakhshia R., Petrovaa D., Pococka R., Imanib M., Seifalian A.M. (2007).
[18]Morris, S., DeYong, C., Wu, Z., Salman, S. and Yemenu, D. DIVA: a visualization system for exploring documents databases for technology forecasting. Computers & Industrial Engineering, Vol. 43 No. 4, pp. 841-62 (2002).
[19]NSTC [National Science and Technology Council] report. The National Nanotechnology Initiative: Research and Development Leading to a Revolution in Technology and Industry (2006).
[20]Purba, N.S. and Nooraeni, Rani. Forecasting of Quantum Dots Technology using Simple Logistic Growth Curve. International Conference on Trade 2019 (ICOT 2019).
[21]Rahmawati, S., Suprijadi, J., Zulhanif. Text mining factor analysis (TFA) in green tea patent data. AIP Conference Proceedings 1827(1):020040 (2017).
[22]Rai, Shivani & Rai, Aditya. Nanotechnology - The secret of fifth industrial revolution and the future of next generation. Nusantara Bioscience. Vol. 7, No.2, pp. 61-66 (2015).
[23]Yoon, J., Choi, S., & Kim, K. Invention property-function network analysis of patents: A case of silicon-based thin film solar cells. Scientometrics, 86(3), 687–703 (2011).

[24]Zhang, J., Wolfram, D., Wolfram, D., Wang, P., Hong, Y., & Gillis, R. Visualization of healthsubject analysis based on query term co-occurrences. Journal of the American Society for Information Science and Technology, 59(12), 1933–1947 (2008).