UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÉNCIAS EXATAS E TECNOLOGIA

Relatórios Técnicos
do Departamento de Informática Aplicada
da UNIRIO
Nº 0002/2022

Exploring the state of the art of Argumentation Mining technology using patent documents

Marcelo Machado
Maria Augusta S. N. Nunes
Rita Pinheiro-Machado
Sean Wolfgand Matsui Siqueira

Departamento de Informática Aplicada

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
Av. Pasteur, 458, Urca - CEP 22290-240
RIO DE JANEIRO - BRASIL

Exploring the state of the art of Argumentation Mining technology using patent documents

Marcelo Machado[1] Maria Augusta S. N. Nunes[1] Rita Pinheiro-Machado[2]
Sean W. M. Siqueira[1]

[1]Programa de Pós-Graduação em Informátic (PPGI) - Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro–RJ, Brazil

[2]Instituto Nacional da Propriedade Industrial (INPI), Rio de Janeiro–RJ, Brazil

marcelo.machado@edu.unirio.br, gutanunes@uniriotec.br, ritap@inpi.gov.br, sean@uniriotec.br

Abstract. Argumentation Mining (AM) has become a hot topic, attracting the interest of research communities ranging from artificial intelligence to computational linguistics, natural language processing, social, and philosophical sciences. Some studies have raised research questions about primary works in the area to understand and support a research agenda. However, no survey has raised questions about the inventions already produced in the area. Unlike other studies, this paper presents a systematic mapping of the technical side of AM using patent documents as a source of information. The systematic protocol led us to the analysis of $N = 50$ documents in which we sought to understand the advancements in AM according to its foundation and gather general information on the area. According to the findings, AM is still an under construction problem with no clear boundaries due to this research domain's novelty. The main inventions are related only to the AM pipeline's initial tasks. Despite that, great progress has been made in projects such as the IBM Project Debater.

Keywords: Patent Documents, Natural Language Processing, Argument Mining, Computational Argumentation.

Contents

1 Introduction

Argumentation is an intelligent communication task that is inherent to human behavior. People argue for various reasons, but to put it in a nutshell, mainly to resolve controversial points. We are very keen on convincing others of our opinion and bringing arguments to a discussion to support our claims. Argumentation plays a fundamental role in decision making on important problems in our society. An example is in the political scenario, where candidates and governments use rhetoric techniques to convince the population about their points of view. It is easy to associate argumentation with political debates, business presentations, or trial court scenarios. However, argumentative elements are also in our daily lives when accessing a social network, reading the news or scientific papers, or even while reading a review of a product and its associated comments.

Finding arguments in an automated way in human discourse was early on discovered as a desirable characteristic of intelligent machines or agents, often referred to as argumentation mining or argument mining [Moens, 2018]. Argumentation mining (AM) is identified as a multidisciplinary research topic, with roots in rhetoric and philosophy, and gained the interest of the scientific community because of its potential when novel Artificial Intelligence (AI) algorithms and techniques are exploited [Lytos et al., 2019]. The recent advances in Machine Learning (ML) combined with the emergence of the social web can enable impressive progress in different scientific fields with a great impact on commercial applications. An AM system deals with the automatic identification of argumentative structures within Natural Language in a great volume of text data through a variety of sources, providing tools for policymaking and Sociopolitical Sciences [Liebeck et al., 2016; Addawood and Bashir, 2016; Boltužić and Šnajder, 2014], Software Engineering [Kurtanović and Maalej, 2018], while it opens new horizons for the broader areas of business, economics, finance, and education.

Argumentation Mining is a broad umbrella for a new set of challenges where different understandings coexist and contribute towards a common yet underspecified objective [Lippi and Torroni, 2015]. However, in recent years, results in this area have been reported, and literature review studies were proposed to investigate the research progress in AM [Lippi and Torroni, 2016; Lytos et al., 2019]. Nevertheless, to the best of our knowledge, none of them set out to collect evidence and present a broader view of the area according to the technical state-of-the-art using patent documents as an input of the analysis.

According to the World Intellectual Property Organization (WIPO)[1], the technological information found in patent documents, which is not often available in any other type of publication, can help researchers and entrepreneurs avoid duplication of efforts and improve existing technologies. The patent documents deposited are associated with identifiers, such as the IPC[2]., which thoroughly facilitates the indexing of technologies, providing access to the state-of-the-art in a specific technological field. Hence, this paper aims to provide a map of the advancement of inventions in AM, and consequently, present an overview of the area. To accomplish this goal, we carried out a systematic mapping of the inventions in AM.

From an initial set containing 710 documents, our conduction process selected 50 of these to extract information such as the main tasks related to AM, the main fields of invention, the timeline of document applications, the main holders and inventors during these years, and the

[1]WIPO is the global forum for intellectual property services, policy, information and cooperation: `https://www.wipo.int/portal/en/index.html`

[2]The acronym IPC stands for International Patent Classification. It is a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain: `https://www.wipo.int/classifications/ipc/en/`

countries or patent offices where the applications occurred. We argue that this survey should help researchers, developers, or anyone interested in understanding the state-of-the-art on this research topic, proposing new artifacts, and developing new tools to be used in the most diverse domains.

The remainder of this paper is organized as follows: Section 2 presents the research method used for this systematic mapping of the literature; Section 3 reports the results of this study according to the research questions; Section 4 discusses the main findings; Section 5 addresses some threats to the validity of our study; finally, Section 6 concludes the paper and presents some perspectives of future work.

2 Methodology

This work follows the systematic mapping of the literature method. It aims to provide an overview of a broad topic by identifying, analyzing, and reporting the available studies[3]. This method seeks to answer a set of research questions defined by the authors. The guidelines used here were based on Silva et al. [2016], Silva et al. [2018], and in the main activities proposed by Petersen et al. [2015]: planning, conducting, and reporting.

2.1 Systematic Mapping Planning

We identified the objectives and defined a protocol to select relevant patent documents during the planning activity. The protocol also aims to reduce researcher bias and make the process reproducible [Steinmacher et al., 2013]. This section summarizes the objectives and the protocol.

2.1.1 Research questions

This work's main objective is to provide an overview of the advances in argumentation mining based on its foundations. Thus, we select secondary research questions to provide evidence for the main question and assist in an overview of the area. Table 1 presents the research questions and brief descriptions of the data to be extracted.

2.1.2 Search string formulation

We started with the terms used to define the area, for instance, "argumentation mining" and "natural language argument", and decided to add terms that involve broader areas, for instance, "computational argumentation" and "natural language processing". Then, to be more specific, we added terms that represent components of argumentation (e.g., "claim" and "evidence") and terms related to argument mining processes (e.g. "rhetorical analysis"). Finally, we put all these selected terms and synonyms together using logical operators to find specific studies in this discussed research topic. Therefore, the final search string was configured, as shown in Table 2.

The final search string is initially composed of an aggregation of terms that define the area. Then, it is aggregated by a composition of atomic components (e.g., "claim" and "evidence") and broader ones (e.g., "NLP" and "text mining"). We understand that this last part may return

[3]"Studies" is the word used by the systematic mapping of the literature method to identify the set of literary studies on a given area, whether they are state of the art papers, state of the art patent documents or other literary artifact defined by the author.

Table 1: Research questions and data to be extracted from patent documents

Research Question	Data to be extracted
Primary Research Question	
PRQ: How advanced are the Argumentation Mining field inventions in line with its foundations?	Type of invention being proposed from the patent description, mainly in the fields of invention and claims. Then, relate this to the argumentation mining tasks and with a broader view of the area in light of the secondary research questions.
Secondary Research Question	
SRQ1: What are the main argumentation mining tasks related to the patent documents?	Argumentation Mining tasks cited in the patent document description.
SRQ2: What are the main IPC codes used to classify this area's patent documents?	The IPC codes field on the cover sheet of the patent document.
SRQ3: In what years were the patent documents filed?	The year in which the patent document was filed.
SRQ4: Who are the patent holders?	The field "applicant" on the patent document's cover sheet. It could be a person, a business entity, or any institution.
SRQ5: In what countries/offices were the patent documents filed?	The country or patent office in which the patent document was filed.

many studies that are not compatible with this research's objective; however, we chose a more generalized search and, in contrast, a more specified filter through the inclusion and exclusion criteria and quality check (presented in Subsection 2.1.4).

2.1.3 Electronic databases

Patentscope[4], which belongs to WIPO, was the electronic database chosen to select the patent documents due to its ability to search for patent documents filed in many offices worldwide and thus be able to analyze what is being globally presented in the area of interest. Patentscope allows us to search 91 million patent documents, including 3.9 million published international patent applications (PCT[5]). Last but not least, Patenscope has the advanced search capability that has allowed us, for instance, to use logical expressions to aggregate and compose the key terms.

The mapping was also carried out on the INPI[6] database to analyze what is being presented in the area of interest in Brazil. However, no study was returned despite we searched separately for small pieces of our final search string.

[4]Patentscope: `https://patentscope.wipo.int/search/pt/search.jsf`

[5]The Patent Cooperation Treaty (PCT) is an international patent law treaty, concluded in 1970. It provides a unified procedure for filing patent applications to protect inventions in each of its contracting states. A patent application filed under the PCT is called an international application, or PCT application.

[6]The acronym INPI stands for National Industrial Property Institute (Instituto Nacional da Propriedade Industrial): `https://www.gov.br/inpi/pt-br`

Table 2: Final search string defined to select relevant patent documents

English version of the final search string
"natural language argument" OR "argument mining" OR "computational argument" OR "debate mining" OR (((claim AND argument AND (evidence OR premise) AND mining) OR "rhetorical analysis") AND ("natural language processing" OR "NLP" OR "text mining"))
Portuguese version of the final search string
"argumento em linguagem natural" OR "mineração de argumento" OR "argumentação computacional" OR "mineração de debate" OR (((afirmação AND argumento AND (evidência OR premissa) AND mineração) OR "análise de retórica") AND ("processamento de linguagem natural" OR "NLP" OR "mineração de texto"))

2.1.4 Selection criteria

Our goal was to include only relevant studies related to AM. The process used to include and exclude a patent document was organized in one inclusion criteria (IC) and one exclusion criteria (EC):

- IC1: The patent document must be related to argumentation mining

- EC1: Duplicated patent documents

We also defined some quality criteria because, despite the exclusion criteria, some documents may not provide the information adequately enough to draw any conclusions. Table 3 presents the defined quality questions and their associated scores.

Table 3: Quality Assessment checklist

Quality Question	**Scores**
QQ1: If the patent body is not written in English or Portuguese. Does the abstract of the patent document present enough information to answer this study's research questions?	Yes (1) N/A (0) No (-2)
QQ2: If the patent body is written in English or Portuguese. Does the abstract of the patent document present enough information to answer this study's research questions?	Yes (1) N/A (0)
QQ3: Was the patent application under the Patent Cooperation Treaty (PCT)?	Yes (1) No (0)
QQ4: Is the patent holder a renowned institution on topics related to argument mining? (e.g., a research institute that presents research on related topics).	Yes (1) No (0)

Our goal with this quality assessment was not to exclude documents but to grade the patents according to our criteria unless they were not suitable to answer our research questions. That is why in QQ1 we associate negative scores to a patent document in which it is impossible to gather

the information we need. Only in this case, patents were excluded because we set a threshold equal to 0 (i.e., $sum(scores) <= 0$).

2.2 Systematic mapping conduction

After the definition of the electronic databases and the search string, the following selection process of the patent documents was applied:

1. Execution of the search string in the selected databases and removal of duplicates after merging the returned results;
2. Analysis process by reading the title and abstract of the patent documents;
3. Data extraction.

It is important to note that this research was conducted in October 2020. Figure 1 depicts the systematic mapping conducting process.

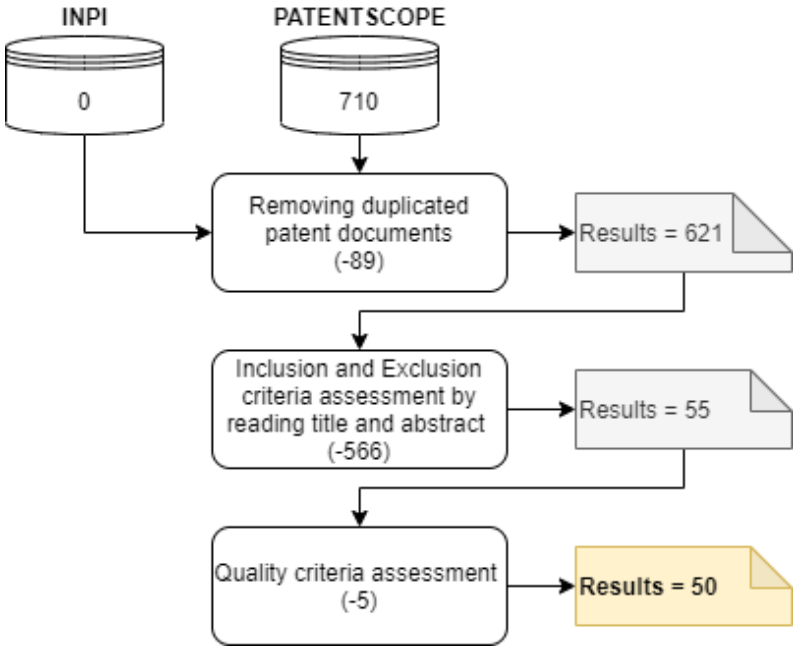

Figure 1: Systematic mapping conducting process

3 Systematic mapping report

We fully analyzed the 50 selected patent documents and extracted the information to answer the research questions (hereafter, we will use PD followed by a number to identify the patent documents, e.g., PD1 – all patent documents are depicted in Table 7 in the Appendix section). This section first presents the systematic mapping results from the secondary research questions' information. Then, we gathered the main results and summarized them in Table 6. The primary research question will be discussed in Section 4 in light of these results.

3.1 SRQ1: What are the main argumentation mining tasks related to the patent documents?

The AM objective is to automatically extract arguments from generic textual corpora and then provide structured data for computational models of argument and reasoning engines [Lippi and Torroni, 2016]. Some studies aim at extracting the arguments from generic unstructured documents, which is a fundamental step in practical applications [Levy et al., 2014]. Others start from a given set of arguments and focus on identifying attack/support relations between the arguments [Boltužić and Šnajder, 2014]. Moreover, the term "argument" sometimes is treated as a composition of internal components such as a claim, evidence (or premise), and a support relation that joins the former ones. Other times, an argument is treated as an atomic structure [Lippi and Torroni, 2016], a basic unit that is not composed of other elements but composes larger structures.

Three main steps are associated with AM: (1) Segmenting a text into argument and non-argument units [Ajjour et al., 2017]; (2) classifying the type of each unit [Rinott et al., 2015]; and (3) identifying support and attack relations between units [Peldszus and Stede, 2013]. To answer this research question, we sought to identify which tasks related to these steps were used in the patent documents.

We identified a set of tasks from the analysis of the patent documents: claim extraction, evidence extraction, opinion extraction, stance classification, polarity extraction, content improvement, machine learning, argument similarity, logic components identification, argument analysis, argument structuring, and argument evaluation. To illustrate some cases, PD7 e PD8 created a method that receives a topic under consideration detecting one or more claims relevant to that topic in the content, and we classify those documents in the claim extraction task; PD26 proposed systems, devices, and methods to detect affective argumentation in a text to support chatbots, and we also classify that document in the claim extraction task.

The identified tasks have similar features or even have different names to represent the same set of actions. Thus, we aggregate these tasks into four groups in which we sought to classify the patent documents, namely: Argumentation Extraction (AE), Argumentation Construction (AC), Argumentation Validation (AV), and Knowledge Representation (KR). Table 4 depicts these groups, along with their description and related tasks.

Figure 2 depicts the patent documents' distribution according to the groups mentioned earlier. Moreover, Table 6 presents this group classification for each patent document.

3.2 SRQ2: What are the main IPC codes used to classify this area's patents?

We found many IPC codes (i.e., 65) among the 50 patent documents selected for analysis in this systematic mapping. Figure 3 depicts the most frequently IPC codes used to classify the patent documents related to AM (i.e., frequency bigger than one).

Among the 65 IPC codes, 62 are from Section G (Physics) and the others from Section H (Electricity). The most frequent Class is G06 (Computing; Calculating or Counting), followed by Class G09 (Educating; Cryptography, Display, Advertising; Seals). The most related subclasses are G06F (Electric Digital Data Processing), G06N (Computer Systems based on Specific Computational Models). Furthermore, according to the most related groups and subgroups, the patents are classified in some of the following fields of invention:

- Natural language analysis (semantic analysis of natural language)

Table 4: Groups to classify patent documents according to the tasks of AM they perform

Group	Description	Related Tasks
AE	This group represents the tasks responsible for capturing components of arguments in a given content.	claim extraction; evidence extraction; opinion extraction; logic components identification; polarity extraction; stance extraction.
KR	This group represents the task responsible for structuring arguments components to allow data management and reasoning about a given content.	argument structuring.
AC	This group represents the tasks responsible for creating arguments about a given topic.	content improvement; machine learning.
AV	This group represents the tasks responsible for using argumentation components to validate information according to a given topic.	argument evaluation; argument similarity.

- Discourse or dialogue representation
- Processing or translation of natural language (natural language analysis; semantic analysis)
- Phrasal analysis, e.g., finite state techniques or chunking
- Query formulation
- Knowledge representation
- Inference methods or devices
- Machine Learning and Learning methods

3.3 SRQ3: In what years were the patent documents filed?

Figure 4 shows the timeline of patents filed in argument mining and related research topics. The first patent (PD4) related to AM was filed in 2003, although the term ”argumentation mining” had not been coined.

3.4 SRQ4: Who are the patent holders?

Mapping the holders and inventors of the area may help to find places and people of interest to guide researchers to find the best forums and partnerships to unite researchers and entrepreneurs with different backgrounds working on the same topic.

We analyzed the field “applicant” on the patent cover sheet to answer this question. It usually indicates a company or institute that inventors are associated with and will hold the patent rights. We found 12 different holders in the field, as shown in Figure 5.

We are also interested in the inventors; hence the field ”inventors” in the patent documents’ cover sheet was also analyzed. We found 83 different inventors involved in the development of patents. Figure 6 depicts the most frequent inventors (i.e., more than one patent application).

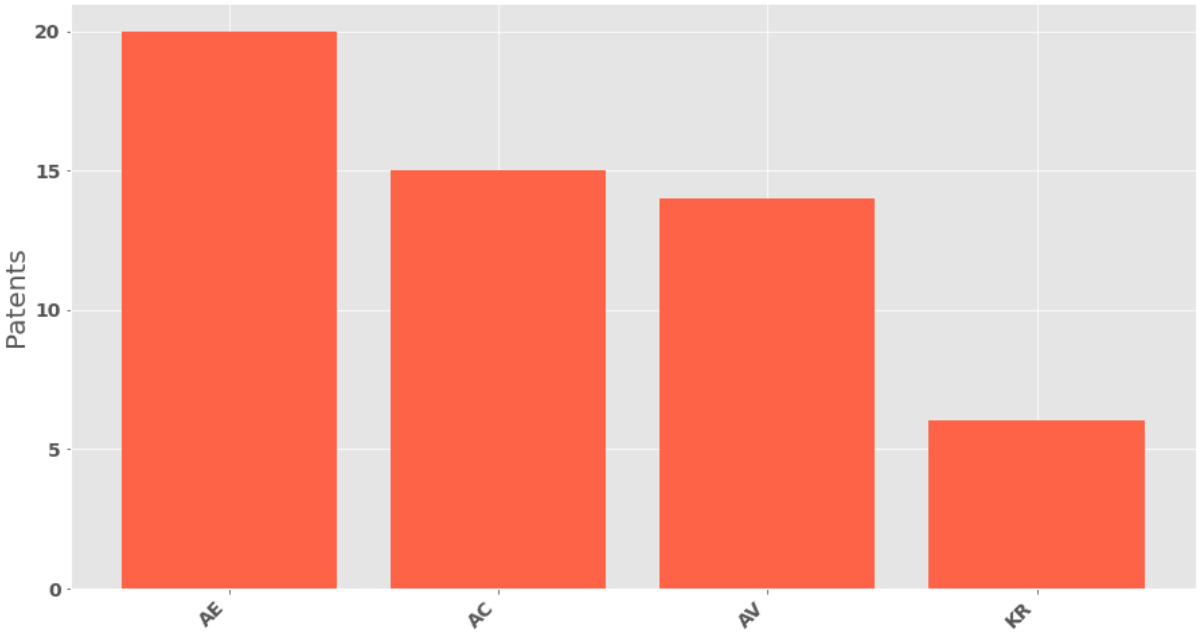

Figure 2: Number of patent documents according to the groups of AM tasks

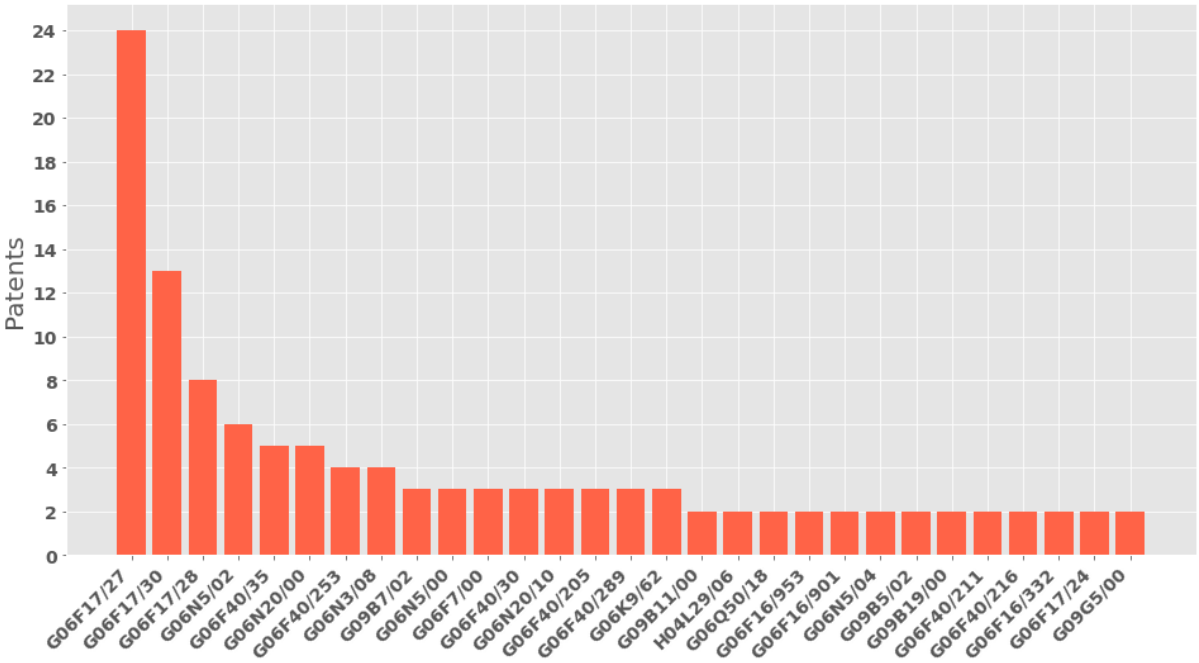

Figure 3: Most IPC codes frequently used to classify patent documents related to AM

3.5 SRQ5: In what country/office were the patent documents filed?

Table 5 depicts the countries/patent offices where our set of analyzed patent documents were filed. The United States of America is by far the country with the most applications. Moreover, seven patent documents were filed via PCT, 6 of which were related to the USA and 1 to Japan.

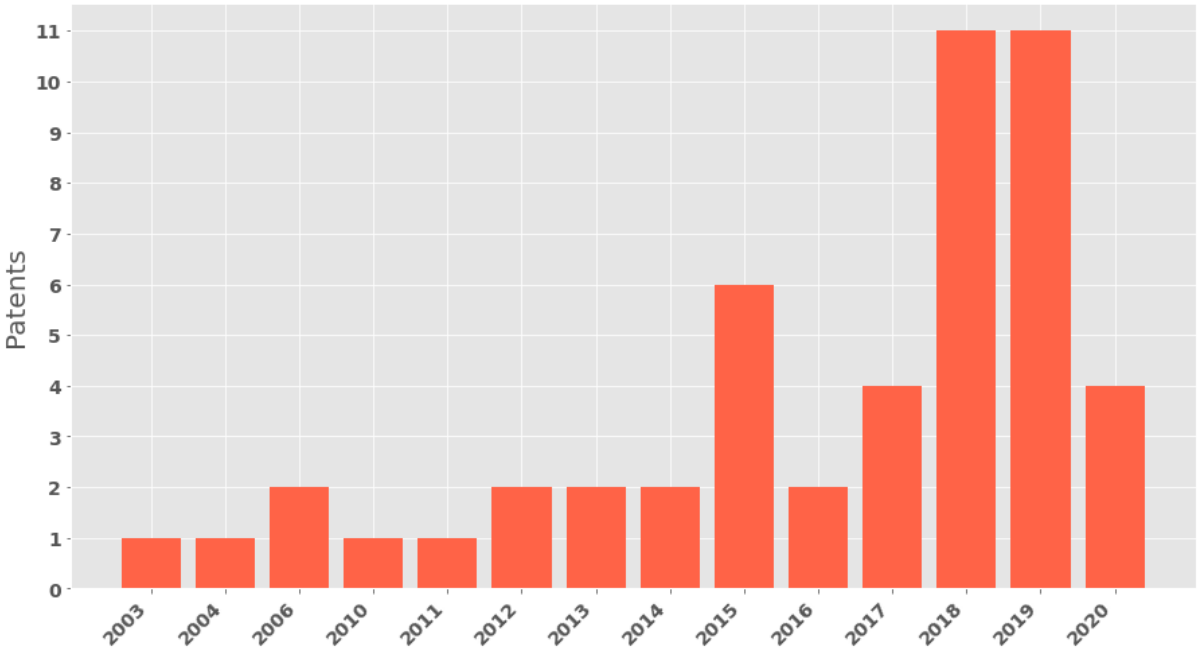

Figure 4: Number of patent documents filed over the years in AM or related topics

Finally, one patent document was filed via EPC[7].

Table 5: Countries/offices that received patent documents application in AM

Country/Office	Total
Japan	1
USA	39
China	2
PCT	7
EPC	1

3.6 Results at a glance

Table 6 summarizes this study's results according to the answers presented in the previous subsections, where ID is the patent document identification, QS the document quality score, C/O country or patent office where the patent was filed, and GT stands for the group of tasks classification.

4 Discussion

Argumentation Mining is an NLP specialization in which the objective is to extract arguments, represent them in a structured way (e.g., in a knowledge base) that allows manipulation for analysis and reasoning. These steps define the foundation of the area. This paper aims to understand how advanced the area's inventions are according to the area's foundation (see PRQ

[7]The European Patent Convention (EPC), also known as the Convention on the Grant of European Patents of 5 October 1973, is a multilateral treaty instituting the European Patent Organisation and providing an autonomous legal system according to which European patents are granted.

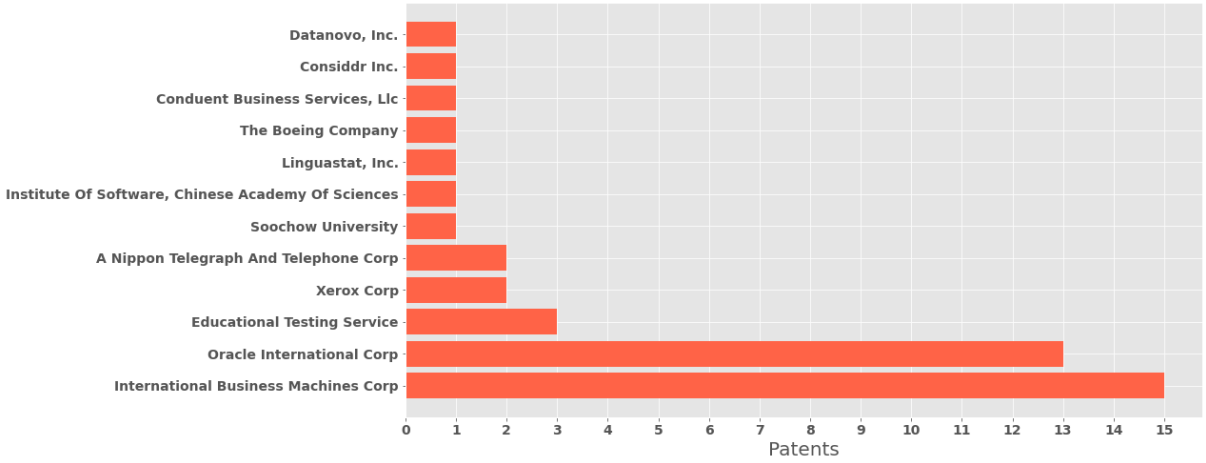

Figure 5: The patent holders in AM and related topics

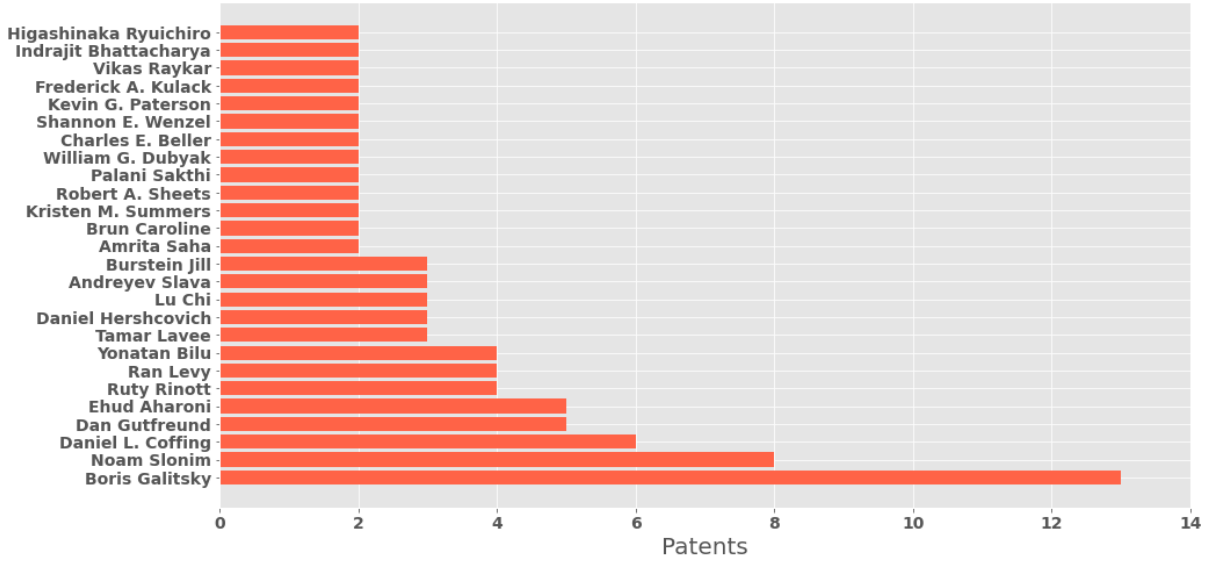

Figure 6: The patent inventors in AM and related topics

in Table 1). Therefore, this section discusses the results aforementioned in Section 3 in line with the primary research question.

Even though the AM foundation is based on big steps from argument extraction to the use of inferences provided by the reasoning capabilities of an argument's knowledge base, some smaller tasks — it does not mean easy — can be extracted from this whole process. In light of our first secondary research question (SRQ1), it is possible to perceive elements that indicate the research area's lack of maturity. According to the tasks of AM, we inferred four groups of patent documents. The results showed that most of the tasks are related to the extraction of arguments, one of the AM process' initial phases. Also, the ability to make inferences (i.e., reasoning) from the arguments recognized in a given content is intended in AM. However, we have not verified documents that address this aspect. This statement can be justified because data needs to be structured for reasoning to allow data management, but, we can verify that the group with the lowest frequency was exactly that of knowledge representation.

It is possible to make some assumptions regarding the inferred groups of tasks and some disciplines, mainly from NLP. For instance, most of the patent documents classified as Argument

Table 6: Summary of the results

ID	QS	C/O	GT	IPC codes	ID	QS	C/O	GT	IPC codes
PD1	2	CN	AE	G06F17/27	PD26	2	US	AE/AC	G06F 17/27; G06N 5/02
PD2	2	CN	AE	G06F17/27	PD27	2	US	AE/AC	G06F40/211; G06F40/30; G06N5/02; G06N20/10; G06N3/00; G06N5/00; G06F40/35; G06F40/205; G06F40/216; G06F40/253; G06F40/289
PD3	3	PCT	KR	G06F16/90; G06F17/28	PD28	2	US	AV	G06F17/27
PD4	2	PCT	AV	G06F17/27; G09B7/02	PD29	2	US	AV	G06F40/211; G06F40/30; G06F40/35; G06F16/31; G06F16/332; G06F40/44; G06F40/51; G06F40/55; G06K9/62; G06N3/08
PD5	1	US	AV	G09B11/00	PD30	2	US	AV	G06F17/27; G06F17/28; G06F17/30; G06K9/62; G06N3/08
PD6	1	US	AV	G09B11/00; G09B7/00	PD31	3	PCT	AV	G06F17/27
PD7	2	US	AE	G06F17/30	PD32	1	US	AV	H04L29/06; G06F16/23; G06F16/11
PD8	2	US	AE	G06F17/27; G06N5/02; G06N5/00	PD33	2	PCT	AV	H04L9/00
PD9	2	US	AE	G10L19/00; G06F17/28; G10L13/027; G06F17/27	PD34	2	US	AC	G06F40/35; G06F40/117; G06F40/253; G06F40/205; G06F40/279; G06F16/953
PD10	2	US	AC	G06F40/00; G06F40/56; G06F40/35; G06N20/00	PD35	2	US	AE	G06F17/27; G06F8/30
PD11	2	US	AV/AE	G06F17/27; G06F17/30	PD36	2	US	AE	G06F17/27
PD12	1	US	AV	G06F17/27; H04L29/06; H04L 12/58	PD37	2	EPC	AE	G06F17/27
PD13	2	PCT	AC	G06F17/28	PD38	2	US	AV	G06F17/30; G06F17/27; G06K9/62; G06N3/08; G06N3/04
PD14	2	US	AC	G06Q10/10; G06Q50/18	PD39	2	US	AC	G06F17/30; G06N5/04
PD15	2	US	AC	G06F17/27; G06N20/00; G06F16/22; G06F16/953; G06N5/00	PD40	1	US	AE	G06F17/27; G06F17/24; G06F17/28
PD16	2	US	AC	G06F16/9032; G06F17/27; G06F17/28; G06F16/901	PD41	2	PCT	AE	G06F17/21; G06F17/27; G10L13/08
PD17	3	PCT	AC	G06F17/27; G06F17/16	PD42	2	US	KR	G06F16/28; G06N20/00; G06F16/901; G06N5/02
PD18	2	US	AE	G06F17/30	PD43	1	US	AE	G06N3/08; G16H70/60; G06 20/10; G06N20/20; G06F16/903; G06Q10/06; G06N5/02; G06N20/00
PD19	1	US	AV/AE	G06F17/30; G06Q50/18	PD44	2	US	AE/KR	G06F17/30; G06F17/27; G06F3/0482
PD20	2	JP	AC	G06N 5/04; G06N20/00	PD45	2	US	AE/KR	G09G5/00; G09G5/00
PD21	2	US	KR	G06F17/30	PD46	2	US	AE	G06F17/24
PD22	2	US	KR	G06F17/30	PD47	2	US	AC	G10L13/04; G10L13/10; G06F16/332; G06F16/35
PD23	2	US	AC/AE	G09B5/02; G09B7/02; G06F17/27; G06F17/28; G09B19/00	PD48	2	US	AC	G06F40/35; G06F40/253; G06 40/289
PD24	2	US	AC	G06F17/28; G06F17/27; G09B5/02; G09B7/02; G09B19/00	PD49	2	US	AC	G06F40/30; G06F40/205; G06F40/216; G06F40/253; G06F40/289; G06N20/10
PD25	1	US	AC	G06F7/00; G06F7/00; G06F7/00; G06F17/30; G06F17/30	PD50	3	US	AV	G06F17/30; G10L25/51; G06F17/27; G06N5/02; G06N99/00

Construction are related to the Question Answering discipline (PD26, PD27, PD47, PD49), mainly with the development of conversational agents such as chatbots. It makes sense since the

purpose of these systems is to respond appropriately to the questions asked, and for that, it is necessary to formulate arguments. Additionally, most of the inventions that deal with Argument Validation tasks are associated with the topic of information trust (i.e., fact check), as we may see in the patent documents PD32, PD33, and PD50.

The IPC codes found (see SRQ2 in Section 3.2), in addition to the answers to the other research questions, confirm how multidisciplinary this research topic is. The answer to the SRQ3 pointed out that the oldest work we selected was from 2003 (PD4). It means that while the term "argument mining" is recent, related topics have been explored since the past decade. Since 2015 it is possible to notice an evolution in patent applications in the area.

According to SRQ4, Boris Galitsky[8] was the most frequent inventor. His name has always been associated with Oracle International Corporation, i.e., one of the leading companies holding patents in this topic. According to Researchgate, his main researched disciplines are AI, Algorithms, and Data Mining. Moreover, he has many papers about Chatbots and Question Answering in general. Following the frequency of inventors, several researchers from IBM Research appear. Then, IBM was the leading company holding document patents related to AM. *IBM Project Debater*[9] is the first AI system that can debate humans on complex topics. Its core is argumentation mining, which justifies the large number of patents filed in that company's name.

Finally, the USA was the main country that received more patent applications. It shows that this research topic has not yet been globally disseminated (see SRQ5 in Section 3.5).

5 Threats to Validity

This systematic mapping aimed to present an overview of the technical state-of-the-art in argumentation mining using patent documents filed over the years. However, there are limitations and threats to this work's validity. Although we have covered 50 patent documents, we do not use other existing electronic databases, threatening conclusion validity. Then, we may have missed some patents in the area. However, we believe that the selected electronic databases were enough to obtain a big picture of the involved area.

There might be bias regarding the number of researchers selecting patent documents in the systematic mapping. The inclusion or exclusion of papers might be subjective or error-prone. However, we mitigated this threat by having more than one researcher checking the inclusion/exclusion and discussing borderline studies.

Our systematic search criteria might also be subject to critique, threatening construct validity. The search string might not contain all relevant keywords, which could cause a loss of some artifacts, and errors could be inserted in the protocol definition. The search string was evaluated using some patent documents to control the results to mitigate this. Patents appeared in the results generating evidence about the search string correctness.

Regarding external validity, with a systematic mapping of the literature, it is important to demonstrate sufficient repeatability. If another set of people analyzes the same group of publications using our set of features, we have confidence that our definitions would help them make choices that are fairly consistent with our results. Although another research group could go through a different process of extracting information, we believe that our results provide a

[8]Boris Galitsky Researchgate profile: `https://www.researchgate.net/profile/Boris_Galitsky`

[9]IBM Project Debater web page: `https://www.research.ibm.com/artificial-intelligence/project-debater/index.html`

useful contribution to the argumentation mining community.

6 Concluding Remarks

This paper presented a systematic mapping of the literature using patent documents related to the argumentation mining research field. Hitherto, no other study in this area organizes this kind of knowledge.

A set of 50 patent documents, from an initial set of 710, was analyzed. We sought to understand the degree of advancement of invention in the area according to the process that underlies it. Besides, we presented an overview of the area in light of our research questions. The results showed that AM is a rather recent innovation field, with most applications coming from the USA. The first patent filed among the documents analyzed was in 2003; however, only after 2015, it is possible to notice an increase in inventions in the area.

Most of the applications were related to companies, with IBM and Oracle the main applicants. IBM is the pioneer on this research topic, and the IBM Project Debater presents impressive results from computers debating complex topics with a human world champion in debates. After gathering the tasks of AM referenced in the patent documents, we inferred four groups to classify these documents, namely, Argument Extraction, Argument Construction, Argument Validation, and Knowledge Representation. It was possible to relate these groups to some domains of the inventions. For instance, Argument Construction was mainly associated with inventions in Question Answering, to be more specific, with the creation of conversational agents such as chatbots. Argument Validation tasks were associated with the topic of information trust; therefore, documents about "Fact Check" and "Fake News" are classified into this group. We believe that as the area evolves, new groups should be inferred.

Finally, we believe that the years ahead will be promising for innovations in this area, given the great advances and natural language processing and machine learning. Besides, the applicability is diverse for the construction of technologies to assist in debates, assess information reliability, provide more effective question answering systems, assist in complex decision-making, encourage critical thinking, just to mention a few examples. As a future work, we intend to perform a systematic review and mapping of the literature on scientific papers to correlate these studies' findings, thus providing a core of evidence of the general state-of-the-art in Argumentation Mining and related topics.

Acknowledgment

This study was financed in part by the 'National Council for Scientific and Technological Development (CNPq) - Brazil' - Process 315374/2018-7, Project 'Searching as Learning: the information search as a tool for learning'. And was financed in part by the Coordination of Superior Level Staff Improvement – Brazil (CAPES) – Finance Code 001.

References

Addawood, A. and Bashir, M. (2016). "what is your evidence?" a study of controversial topics on social media. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 1–11.

Ajjour, Y., Chen, W.-F., Kiesel, J., Wachsmuth, H., and Stein, B. (2017). Unit segmentation of argumentative texts. In *Proceedings of the 4th Workshop on Argument Mining*, pages 118–128.

Boltužić, F. and Šnajder, J. (2014). Back up your stance: Recognizing arguments in online discussions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 49–58.

Kurtanović, Z. and Maalej, W. (2018). On user rationale in software engineering. *Requirements Engineering*, 23(3):357–379.

Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., and Slonim, N. (2014). Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500.

Liebeck, M., Esau, K., and Conrad, S. (2016). What to do with an airport? mining arguments in the german online participation project tempelhofer feld. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 144–153.

Lippi, M. and Torroni, P. (2015). Argument mining: A machine learning perspective. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 163–176. Springer.

Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2):1–25.

Lytos, A., Lagkas, T., Sarigiannidis, P., and Bontcheva, K. (2019). The evolution of argumentation mining: From models to social media and emerging tools. *Information Processing & Management*, 56(6):102055.

Moens, M.-F. (2018). Argumentation mining: How can a machine acquire common sense and world knowledge? *Argument & Computation*, 9(1):1–14.

Peldszus, A. and Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.

Rinott, R., Dankin, L., Alzate, C., Khapra, M. M., Aharoni, E., and Slonim, N. (2015). Show me your evidence-an automatic method for context dependent evidence detection. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 440–450.

Silva, Í. D., Nunes, M. A. S. N., Moro, M. M., and Barbosa, A. S. (2018). Gibis da área da ciência da computação. metodologia científica e tecnológica. *Almanaque para popularização de ciência da computação. Série 6*, 1.

Silva, Í. D., Nunes, M. A. S. N., Pinheiro-Machado, R., and Santos, J. C. (2016). Gibis da área da ciência da computação. propriedade intelectual. *Almanaque para popularização de ciência da computação. Série 3*, 1.

Steinmacher, I., Chaves, A. P., and Gerosa, M. A. (2013). Awareness support in distributed software development: A systematic review and mapping of the literature. *Computer Supported Cooperative Work (CSCW)*, 22(2-3):113–158.

Appendix A: List of the Patent Documents

Table 7 summarizes the patent documents analyzed in this research study.

Table 7: List of the 50 patent documents selected and analyzed in this study

ID	Patent Citation
PD1	Mingxue, L.; Xueyu, D.; Feng, T. An unsupervised thesis point extraction method based on debate mining. Applicants: Institute of Software, Chinese Academy of Sciences. CN109902284. Application Date: 30.12.2018.
PD2	Fang, K; Peifeng, L.; Yuhua, Z.; Guodong, Z.; Qiaoming, Z. Argument extraction method and system. Applicants: Soochow University. CN103530281. Application Date 15.10.2013. Grant Date: 22.06.2016.
PD3	MITSUDA, K.; HIGASHINAKA, R.; TOMITA, J. ARGUMENTATION STRUCTURE EXTENSION DEVICE, ARGUMENTATION STRUCTURE EXTENSION METHOD, PROGRAM AND DATA STRUCTURE. Applicants: A Nippon Telegraph and Telephone Corporation. WO2019160098. Application Date: 15.02.2019.
PD4	Jill, B; Slava, A; Chi, L. AUTOMATED ESSAY SCORING. Applicants: Educational Testing Service. WO2004061593. Application Date: 15.12.2003 .
PD5	Jill, B; Slava, A; Chi, L. Automated essay scoring. Applicants: Educational Testing Service. US20060286540. Application Date: 27.07.2006. Grant Date: 03.08.2010.
PD6	Jill, B; Slava, A; Chi, L. Automated essay scoring. Applicants: Educational Testing Service. US20100297596. Application Date: 30.07.2010. Grant Date: 18.03.2013.
PD7	Aharoni, E.; Bilu, Y; Gutfreund, D.; Hershcovich, D.; Lavee, T.; Levy, R.; Rinott, R.; Slonim, N. Automatic detection of claims with respect to a topic. Applicants: International Business Machines Corp. US20160321336. Application Date: 28.04.2015. Grant Date: 03.07.2018.
PD8	Aharoni, E.; Bilu, Y; Gutfreund, D.; Hershcovich, D.; Lavee, T.; Levy, R.; Rinott, R.; Slonim, N. AUTOMATIC DETECTION OF CLAIMS WITH RESPECT TO A TOPIC. Applicant: International Business Machines Corp. US20190220515. Application Date: 16.01.2018.
PD9	Aharoni, E.; Bhattacharya, I.; Bilu, Y.; Gutfreund, D.; Hershcovich, D.; Raykar, V.; Rinott, R; Shantanu, G.; Slonim, N. Automatic generation of a speech by processing raw claims to a set of arguments. Applicant: International Business Machines Corp. US20150371651. Application Date: 29.04.2015. Grant Date: 05.09.2017.
PD10	Bilu, Y; Levy, R.; Slonim, N. Claim generation. Applicant: International Business Machines Corp. US20180012127. Application Date: 11.07.2016. Grant Date: 15.09.2020.
PD11	Aharoni, E.; Bar-Haim, R.; Bhattacharya, I.; Dinuzzo, F.; Gutfreund, D.; Saha, A.; Slonim, .N; Yanover, C. Claim polarity identification. Applicant: International Business Machines Corp.US20160350278. Application Date: 26.05.2015. Grant Date: 25.04.2017.
PD12	Marcia, N; Dowell, M.; Nixon, T. COMPUTATIONAL LINGUISTIC ANALYSIS OF LEARNERS' DISCOURSE IN COMPUTER-MEDIATED GROUP LEARNING ENVIRONMENTS. US20190138597. Application Date: 30.07.2018.
PD13	Coffing, D.L. COMPUTER-BASED METHOD AND SYSTEM OF ANALYZING, EDITING AND IMPROVING CONTENT. WO2014146086. Application Date: 18.03.2014.
PD14	Coffing, D.L. COMPUTER-BASED METHOD AND SYSTEM OF ANALYZING, EDITING AND IMPROVING CONTENT. US20190392393. Application Date: 27.08.2019.

PD15	Galitsky, B. CONSTRUCTING CONCLUSIVE ANSWERS FOR AUTONOMOUS AGENTS. Applicants: Oracle International Corp. US20200117709. Application Date: 16.10.2019.
PD16	Galitsky, B. CONSTRUCTING IMAGINARY DISCOURSE TREES TO IMPROVE ANSWERING CONVERGENT QUESTIONS. Applicants: Oracle International Corp. US2019034729. Application Date: 09.05.2019.
PD17	Galitsky, B. CONSTRUCTING IMAGINARY DISCOURSE TREES TO IMPROVE ANSWERING CONVERGENT QUESTIONS. Applicants: Oracle International Corp. WO2019217722. Application Date: 09.05.2019.
PD18	Aharoni, E.; Dankin, L; Gutfreund, D.; Lavee, T.; Levy, R.; Rinott, R. Slonim, N. Context-dependent evidence detection. Applicant: International Business Machines Corp. US20160350410. Application Date: 25.05.2015. Grant Date: 03.07.2018.
PD19	Chan, A. H. Data mining and analysis system and method for legal documents. Applicants: Alex Chan, A. H.; DataNovo, Inc. US10133791 . Application Date: 08.09.2015. Grant Date: 20.11.2018.
PD20	Ryuichiro, H.; Hiroaki, S.; Hiromi, N.; Takaaki, F.; Yoshihiro, M. DIALOG SYSTEM, METHOD, AND PROGRAM. Applicants: Nippon Telegr & Teleph Corp. JP2018147189. Application Date: 03.03.2017. Grant Date: 27.03.2020.
PD21	Kulack, F. A.; Paterson, K. G.; Wenzel, S. E. Displaying logical statement relationships between diverse documents in a research domain. Applicants: Kulack, F. A.; International Business Machines Corp.; Paterson, K. G.; Wenzel, S. E. US20120221583. Application Date: 25.02.2011. Grant Date: 14.03.2017.
PD22	Kulack, F. A.; Paterson, K. G.; Wenzel, S. E. Displaying logical statement relationships between diverse documents in a research domain. Applicants: International Business Machines Corp. US20130097191. Application Date: 04.12.2012. Grant Date: 16.05.2017.
PD23	Beller C. E.; Dubyak, W. G.; Sakthi, P.; Sheets, R. A.; Summers, K. M. Document preparation with argumentation support from a deep question answering system. Applicants: International Business Machines Corp. US20190163745. Application Date: 30.11.2017. Grant Date: 20.08.2019.
PD24	Beller C. E.; Dubyak, W. G.; Sakthi, P.; Sheets, R. A.; Summers, K. M. DOCUMENT PREPARATION WITH ARGUMENTATION SUPPORT FROM A DEEP QUESTION ANSWERING SYSTEM. Applicants: International Business Machines Corp. US20190318001. Application Date: 27.06.2019.
PD25	Dyke, P. H. V.; Peter, W.; Sven, B.; John, S. Dynamic information extraction with self-organizing evidence construction. US20050154701. Application Date: 01.12.2004.
PD26	Galitsky, B. ENABLING CHATBOTS BY DETECTING AND SUPPORTING AFFECTIVE ARGUMENTATION. Applicants: Oracle International Corp. US20190138595. Application Date: 28.09.2018.
PD27	Galitsky, B. Enabling chatbots by detecting and supporting argumentation. Applicants: Oracle International Corp. US20180357220. Application Date: 15.06.2018. Grant Date: 09.06.2020.
PD28	Galitsky, B. ENABLING CHATBOTS BY VALIDATING ARGUMENTATION. Applicants: Oracle International Corp. US20190272323. Application Date: 29.01.2019.
PD29	Galitsky, B. Enabling rhetorical analysis via the use of communicative discourse trees. Applicants: Oracle International Corp. US20180329879. Application Date: 09.05.2018. Grant Date: 06.10.2020.

PD30	Galitsky, B. ENABLING RHETORICAL ANALYSIS VIA THE USE OF COMMUNICATIVE DISCOURSE TREES. Applicants: Oracle International Corp. US20180329880. Application Date: 09.05.2018.
PD31	Galitsky, B. ENABLING RHETORICAL ANALYSIS VIA THE USE OF COMMUNICATIVE DISCOURSE TREES. Applicants: Oracle International Corp. WO2018208979. Application Date: 09.05.2018.
PD32	Coffing, D.L. FACT MANAGEMENT SYSTEM. Applicants: Daniel L. Coffing. US20200092301. Application Date: 16.09.2019.
PD33	Coffing, D.L. FACT MANAGEMENT SYSTEM. Applicants: Daniel L. Coffing. WO2020056409. Application Date: 16.09.2019.
PD34	Galitsky, B. GENERATING DESIRED DISCOURSE STRUCTURE FROM AN ARBITRARY TEXT. Applicants: Oracle International Corp. US20200184155. Application Date: 13.02.2020.
PD35	Carter, B. L.; Cole, K. A.; Gandikota, V.; Hsu, J.M. Identifying logic problems in text using a statistical approach and natural language processing. Applicants: International Business Machines Corp. US20170220554. Application Date: 03.02.2016. Grant Date: 17.03.2020.
PD36	Caroline, B. LEARNING OPINION-RELATED PATTERNS FOR CONTEXTUAL AND DOMAIN-DEPENDENT OPINION DETECTION. Applicants: Caroline, B.; XEROX CORP. US20140067370. Application Date: 31.08.2012.
PD37	Caroline, B. Learning opinion-related patterns for contextual and domain-dependent opinion detection. Applicants: XEROX CORP. EP2711849. Application Date: 23.08.2013.
PD38	Aharonov, R.; Dor, L. E.; Halfon, A; Mass, Y; Shnayderman, I.; Slonim, N.; Venezian, L. LEARNING THEMATIC SIMILARITY METRIC FROM ARTICLE TEXT UNITS. Applicant: International Business Machines Corp. US20200125673. Application Date: 23.10.2018.
PD39	Finberg, N.; Christiansen, M.; Pietraho, T. METHOD OF DATA ORGANIZATION AND DATA SEARCHING FOR USE IN CONSTRUCTING EVIDENCE-BASED BELIEFS. Applicants: Considdr Inc. US20180365324. Application Date: 22.08.2018.
PD40	Coffing, D.L. PROCESSING NATURAL LANGUAGE ARGUMENTS AND PROPOSITIONS. Applicants: Coffing, D. L. US20190370335. Application Date: 19.03.2019.
PD41	Coffing, D.L. PROCESSING NATURAL LANGUAGE ARGUMENTS AND PROPOSITIONS. Applicants: Coffing, D. L. WO2019183144. Application Date: 19.03.2019.
PD42	Khapra, M.; Raykar, V.; Saha, A.; Slonim, N.; Verma, A. Semantic merge of arguments. Applicant: International Business Machines Corp. US20150370887. Application Date: 29.04.2015. Grant Date: 07.04.2020.
PD43	Sen, A.; Mannarswamy, S.; Sinha, M.; Roy, S. Stance classification of multiperspective consumer health information. Applicants: Conduent Business Services, LLC. US20180218253. Application Date: 31.01.2017. Grant Date: 21.04.2020.
PD44	Gordon, M. S.; Kozloski, J. R.; Lenchner, J.; Pickover, C. A. SYSTEM AND METHOD FOR AUTOMATICALLY PROVIDING ALTERNATIVE POINTS OF VIEW FOR MULTIMEDIA CONTENT. Applicants: International Business Machines Corp. US20190163792. Application Date: 27.11.2017.
PD45	Oscar, K. System, method, and computer program product for anticipatory hypothesis-driven text retrieval and argumentation tools for strategic decision support. Applicants: The Boeing Company. US20070018953. Application Date: 27.06.2006.

PD46	Hellwig, J.; Pierre, J. M.; Butler, M. H. SYSTEMS AND METHODS FOR IDENTIFYING CLAIMS IN ELECTRONIC TEXT. Applicants: Linguastat, Inc. US20140189485. Application Date: 07.03.2014.
PD47	Galitsky, B. USING COMMUNICATIVE DISCOURSE TREES TO CREATE A VIRTUAL PERSUASIVE DIALOGUE. Applicants: Oracle International Corp. US20200286463. Application Date: 06.04.2020.
PD48	Galitsky, B. USING COMMUNICATIVE DISCOURSE TREES TO DETECT DISTRIBUTED INCOMPETENCE. Applicants: Oracle International Corp. US20200265195. Application Date: 18.03.2020.
PD49	Galitsky, B. Utilizing discourse structure of noisy user-generated content for chatbot learning. Applicants: Oracle International Corp. US20200218859. Application Date: 07.01.2020.
PD50	Can A. E.; Bull, B.; Carrier, S. R.; Mansjur, D. S. VALIDATING BELIEF STATES OF AN AI SYSTEM BY SENTIMENT ANALYSIS AND CONTROVERSY DETECTION. Applicants: International Business Machine Corp. US20190370391. Application Date: 05.06.2018.