Concept Timestamping on Blockchain and Decentralization of Patents

Kar Seng Loke(✉)

Data Smart PLT, 58100 Kuala Lumpur, Malaysia
Loke.ks@gmail.com

Abstract. We describe a system for registering new inventions and design based on conceptual graphs and the blockchain. Instead of describing the innovation in text as in patent claims, the claims are documented using conceptual graphs as a structured graphical method for capturing concepts and their relationship. Conceptual graphs as a structured graphical method allows easier automatic semantic comparisons. It obviates the need from extracting semantic concepts from text using natural language processing techniques. The graphs then can be hashed and submitted to the blockchain for timestamping to claim the innovation.

Keywords: Blockchain · Patents · Timestamping · Conceptual graphs · Concept timestamping · Decentralized systems

1 Introduction

Intellectual property (IP) refers to creations of the mind that includes inventions, literary and artistic works according to World Intellectual Property Organization (WIPO). Protection to the creators for these intellectual goods are provided by IP laws so that economic benefits may accrue to them while at the same time taking into consideration the wider public interest so that innovation can flourish as well. IP protection includes patents, copyrights and trademarks. Patents, a dominant system for IP, encourage research and development (R & D) and at the same time allow diffusion of knowledge. It is intended to solve the issue of economic incentives and knowledge diffusion. A patent gives the inventor the right to decide who can use his invention. In exchange for this right, the inventor has to disclose the technical information for it.

Patent assessment is a key component in the patent system. Patents are awarded based on novelty and the non-obviousness of the invention. The increase in the number of patent application creates additional demands on the patent examiners because of the difficulty of keeping pace with the increase in workload and increasing technological complexity [1]. This creates difficulty in determining if a claimed invention is really novel and non-obvious. Good quality patents are said to promote innovation and economic benefits. Dubious and low-quality patents are those that have a lower threshold of non-obviousness or novelty. They can result from improper examinations and assessments. This can lead to stifling of innovation and greater uncertainty that may lead to disputes that has be determined by legal means.

G. Wang et al. (Eds.): iSCI 2019, CCIS 1122, pp. 1–10, 2019.
https://doi.org/10.1007/978-981-15-1301-5_50

One suggestion to improve the quality of patents is to expand the number of independent reviewers [2, 3] and by crowdsourcing. Another approach may be to employ computational methods to aid in determine prior art and non-obviousness of the invention.

Patent claims are semi-structured text [4]. The description section is divided into claims, diagrams and description of invention. Computational patent text improvement approaches can be classified into 3 general approaches:

1. Perform structural analysis of text and highlighting important section of texts using natural language [5–7].
2. Improving searching, indexing and retrieval mechanism [8]. This could also involve the creation of structured databases so that comparisons can be performed easier.
3. Use of automatic classification and mining [4] for patterns within text. Vector space methods have been used to classify texts, using machine learning techniques such as Latent Semantic Analysis (LSA), the semantic difference (distance) between patent text can be determined.

There are still challenges remaining for making patents more accessible. We would like to make a preliminary proposal to use a supplementary system for recording innovation and novel inventions that may ameliorate some of the shortcomings of the patent system in relation to the ability to compare inventions and prior art. We propose the use conceptual graphs (CG) to describe the invention and using CG for determining prior art and novelty automatically through graph matching algorithms. The blockchain is used for decentralized timestamping the invention after it has been determined to be sufficiently novel. The entire process again can be automated without a central authority.

2 Related Works - Conceptual Graphs

A conceptual graph (CG) [9] is a graph-based method for representing structural and semantic relationships between concepts. CG is based on the ideas of semantic networks [10] and existential graphs [11]. The concept nodes can represent entities, events, actions, states, attributes etc. A CG consist of a network of concept nodes linked by relation nodes. Mineau et al. [12] have argued that CG through its simplicity and graphical constructs, it is simple to use and yet sufficiently expressive. Graphs such as CG have non-linear structures that are make easy to identify similarities and differences.

For example, in the example below Stock and 10 are concepts and quantity (Qty) relationship links them up. The rectangular box represents concepts and circular shape gives the relationship (rel) between the concepts. The CG can be read as: the quantity (qty) of the stock is 10 (Fig. 1).

AQ1

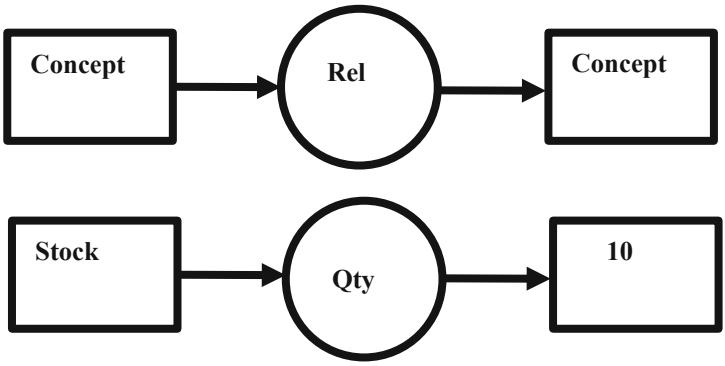

Fig. 1. Example of Concept Graph (CG)

For example, the sentence: "John is going to Boston by bus" can be represented in a conceptual graph (CG) as:

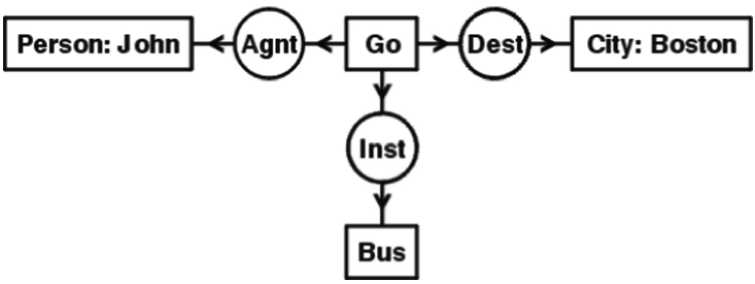

Fig. 2. Example of CG.

In Fig. 2 two of the concepts have names which represents the referent John or Boston. The type of relations is agent (Agnt), destination (Dest), or instrument (Inst). The CG has the following meaning: John is the agent of the concept Go which has destination City that is Boston, and a bus is the instrument.

CG have been used in a variety of domains for knowledge representation, query processing and automated extraction of knowledge. In the medical domain, [13] have described CG representation and querying system to improve tumor classification. SNOMED [14] also uses a manual input CG system for describing patient-description (e.g. headache, cough, chest pain) with modifiers (severity, duration, location, etc.). Another CG input system [15] are able to generate text description based on CG input. Image retrieving based on CG have also been described [16, 17]. CG have also been used in assisting of writing software requirements. Ryan and Mathews [18] described a CG system for documenting software requirements using a domain language. The tool developed allowed previous requirement definitions to be retrieved so that it can be reused. They described a CG comparison tool to identify likely CG for re-use. It recognizes that a similar problem has been solved before. However, it is a text-based tool using the textual form to write or input CG. Business processes documentation has also

been applied using CG. In their comparison, Gerbé et al. [19] have found that CG is more useful when compared to other formalism such as UML (Unified Modeling Language), PIF (Process interchange format) and WfMC (Workflow Management Coalition).

There have some works on CG related to patents, particularly related to patent summarization using natural language processing (NLP). Yang and Soo [20] described a system using a NLP parser that creates CG that summarizes patent claims. It can also be used for comparing patent claims using CG graph matching algorithm [21]. In [22], Rao et al. used conditional random fields to identify the associated relationships expressed concepts in patent claims, these are then used to construct the CG. Rao and Devi [23] also described a system that parses text to extract CG from sentences in the claim section of patent text using Deep Boltzmann Machine. Relationships are extracted based on linguistic rules, for example if concept c1 and c2 are connected by "is", then the relationship is a "sub-type" or "isa" relationship. The issue with extracting CG from text is that it tends to follow the sentential form rather closely and may obscure the conceptual form.

3 Our Approach

Trusted timestamping and proof of existence of documents have been numerously proposed on the Blockchain. Through cryptographic protocols and the irreversibility of the blockchain, document cryptographic hash inserted into the blockchain provides sufficient proof of the document existence at a particular point in time. Moreover, the holder of the private key can claim ownership of the document if it can be shown that the document owner has at all time control of his private key (Fig. 3).

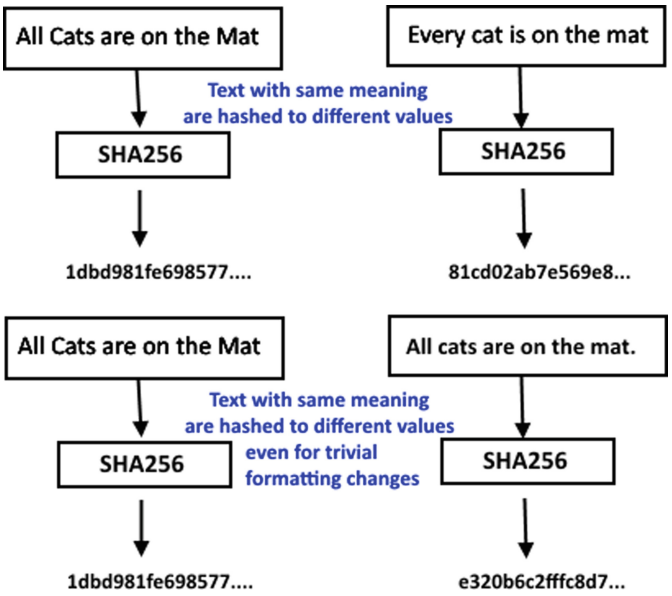

Fig. 3. Text with same meaning will hash to different values

However, documents do not equal content. Two documents with same semantical content except that one has an extra space tagged at the end, will produce a different cryptographic hash. A cryptographic hash is also useless in the case of images as a series of the same photos of different scales (size) will produce different hash value. We can't determine from the hash that they are the same photo but off different sizes. Similarly, a textual document with variations in formatting, capitalization, etc. but of the same semantical content will produce a different hash. We therefore cannot use the hash to determine similarity of content but of document. A hash value only provides proof of a particular document and does not rule out the existence another version of similar content (Fig. 4).

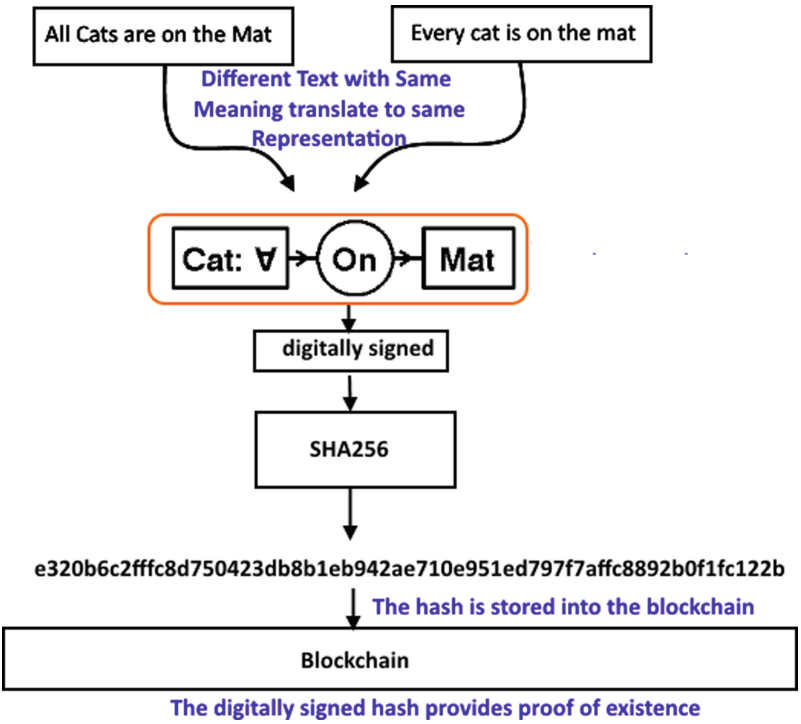

Fig. 4. Concept timestamping and registration

Language (and text) itself is very expressive and can express the same semantical contents with an infinite variety of ways. We seek an approach that is less flexible and more economical in its expression so that different concept will have different expressions, but similar concept should not differ in their expression. Once we have this expression, we can obtain its cryptographical hash to represent the concepts. This hash will then be embedded into the blockchain and claims of its existence of the concept at a particular time can be made.

Languages and formalism for expressing concepts have been developed in the area knowledge representation under cognitive science and artificial intelligence. Various suitable formalism includes First order Logic, Description Logic [24] and Conceptual Graphs [9]. Ideas and concepts then can be expressed in natural language and converted automatically to Conceptual Graphs structure. These structures would be more concise. Yang and Soo [20] have worked on extracted CG from patent text. We propose to work the other way, that is, the patentable concepts are themselves written in text and translated to CG. These CG are then held as canonical representation of those concepts. There are advantages in using the hash of the CG instead of text. As mentioned earlier various text can represent exact same semantical content and will generate different hashes. So, the existence of a hash for a document does not rule other possible text that has the same content. On the other hand, concepts represented by CG should not differ if they are indeed the same concept, so we would have a unique representation of it as the hash. So, if the idea in the CG creates a collision in the hash, then we would know that the idea is not unique and should be rejected by the system.

The centralization of patents through a governmental entity can be done away by incorporating the canonical CG and registering them on the Blockchain. This can be done by obtaining the cryptographic hash of the CG after the authors have signed it with their private keys. The cryptographic hash of the CG serves as digital key of their ideas and concepts. Because of the irreversibility of the blockchain, the timestamp of in the block header would serve as a timestamping measure for their authorship.

4 Proof of Concept

Consider this partial patent claim (US Patent 6103979):

> *"What is claimed is:*
> *1. A keyboard comprising:*
> *a plurality of keys, each key having a character illustrated at each end, and each key having an elongated shape elongated in a first direction, said plurality of keys being aligned along a second direction different from said first direction, thereby forming a plurality of horizontal rows of characters along said second direction; and first and second mutually different electric contacts provided for each key of said plurality of keys in respective, mutually different locations, said key selectively establishing one of said two electrical contacts as a result of being pressed appropriately, wherein a front surface of each key of said plurality of keys is parallelogram in shape, wherein said first and second directions form an angle other than 90°, wherein each of said parallelogram-shaped keys has a longitudinal axis thereof extending along said first direction, said first and second mutually different electric contacts being located beneath two respective ends of each said key along said longitudinal axis thereof."*

We can create a partial CG in the figure below using the CharGer software [25]:

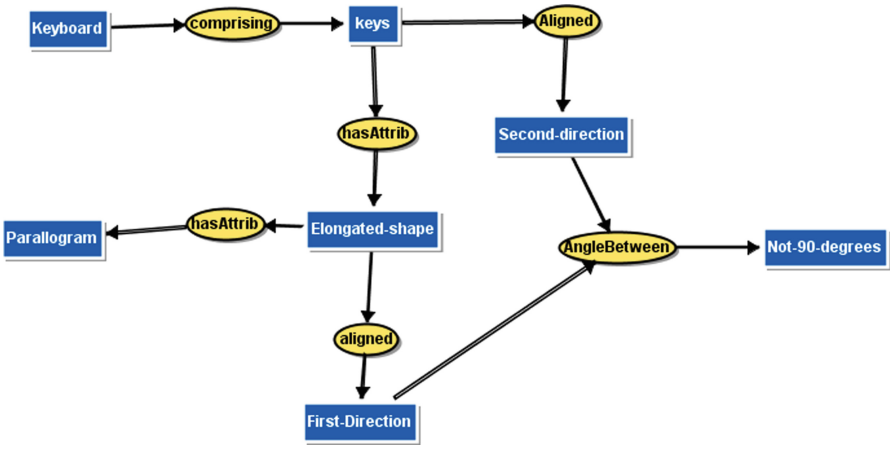

Fig. 5. Partial CG for the patent claim US Patent 6103979

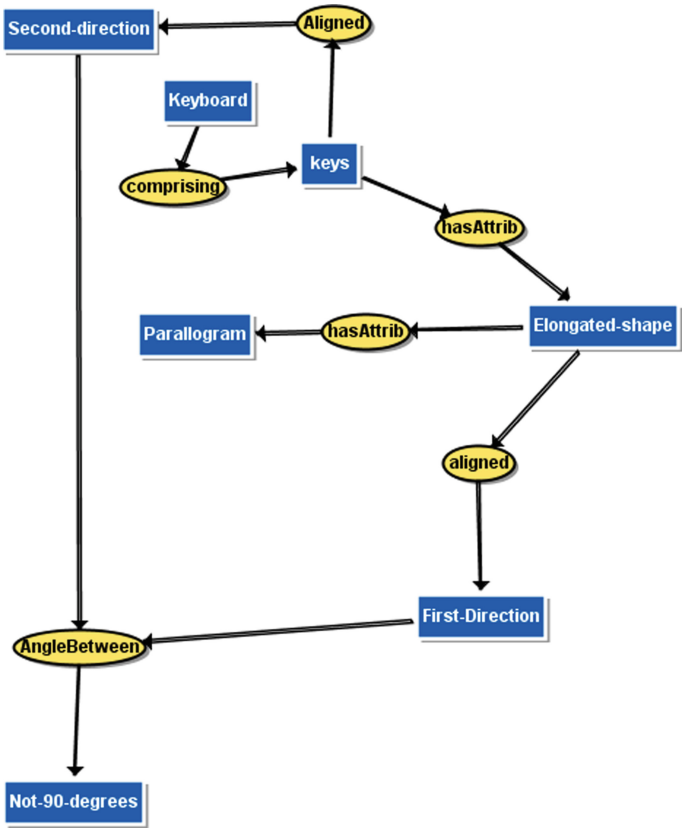

Fig. 6. Partial CG for the patent claim US Patent 6103979 with a different spatial arrangement

Both diagrams (Figs. 5 and 6) translated to linear CGIF (CG interchange format) textual form would have the same content:

[Not-90-degrees: *x1] [Second-direction: *x2] [First-Direction: *x3] [Elongated-shape: *x4] [Keyboard: *x5] [keys: *x6] [Parallogram: *x7] (comprising ?x5 ?x6)(hasAttrib ?x6 ?x4)(aligned ?x4 ?x3)(hasAttrib ?x4 ?x7)(Aligned ?x6 ?x2)(AngleBetween ?x3 ?x2 ?x1)

The CGIF format of the CG would be digitally signed and hashed and submitted to the blockchain for timestamping. Consider another CG derived from another patent claim description [22] in Fig. 7.

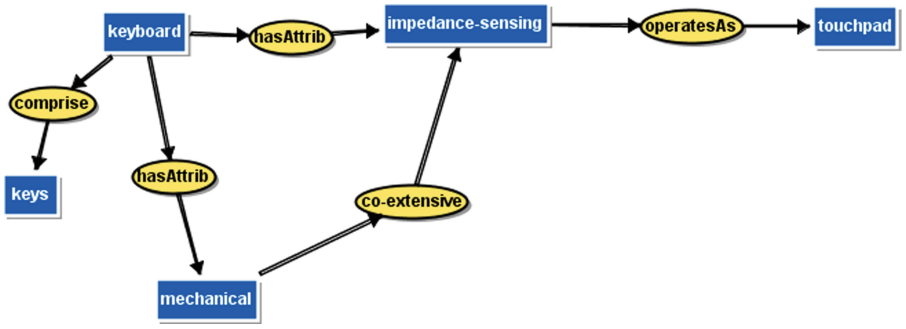

Fig. 7. CG (partial) for a patent claim description

In the case for the CG in Fig. 7, the CGIF is:

[touchpad: *x1] [impedance-sensing: *x2] [mechanical: *x3] [keys: *x4] [keyboard: *x5] (hasAttrib ?x5 ?x2)(comprise ?x5 ?x4)(hasAttrib ?x5 ?x3)(operatesAs ?x2 ?x1)(co-extensive ?x3 ?x2)

The above 2 CGIF contents can now be timestamped. For this proof of concept, we simply used the OriginStamp document timestamping services [26]. The hash (SHA-256) for the above CGIF is obtained respectively as:

- 4f6145729e381f6680f953de3aa47b01d017778b51672b0336774f928f2cc144
- 7217aab5e333cc6d00ef78d13ba3d125720158fa96cb91c559ca08547787eb10

The hashes can be stored in any blockchain, as an example the hash of one of the CGIF above is given as:

- 7217aab5e333cc6d00ef78d13ba3d125720158fa96cb91c559ca08547787eb10

It is stored as an Ethereum Smart Contract with the following address:

- 0xa66a943be0c956ae90eae98b9703b98a7bbd4309

The entire system would, however, require an off-chain database for storing the original CGIF files. This would be elaborated in further research.

5 Conclusion

In summary we propose the use of CG as a representation of ideas and concepts for the purpose of making a claim of priority. Secondly, we have proposed a low-cost, universal, borderless and decentralized system for filing priority that obviates the use of any centralization through the use of the Blockchain. In addition, the uniqueness of the idea can be checked automatically using semantic graph matching algorithms.

References

1. Levin, R.: A patent system for the 21st century. Issues Sci. Technol. **4**(20), 49–54 (2004)
2. Whalen, R.: Complex innovation and the patent office. Chicago-Kent J. Intellect. Property **17** (1), 226–278 (2018)
3. Wagner, R.P.: Understanding patent quality mechanisms. Univ. Pennsylvania Law Rev. **157**, 2135 (2009)
4. Zhang, L., Li, L., Li, T.: Patent mining: a survey. ACM SIGKDD Explor. Newsl. **16**, 1–19 (2014)
5. Wang, J., Lu, W.F., Loh, H.T.: A two-level parser for patent claim parsing. Adv. Eng. Inform. **29**, 431–439 (2015)
6. Sheremetyeva, S.: Automatic text simplification for handling intellectual property. In: Proceedings of the Workshop on Automatic Text Simplification: Methods and Applications in the Multilingual Society, Dublin, Ireland (2014)
7. Shinmori, A., Okumura, M., Marukawa, Y., Iwayama, M.: Patent claim processing for readability: structure analysis and term explanation. In: Proceedings of the ACL-2003 Workshop on Patent Corpus Processing, Sapporo, Japan (2003)
8. Lupu, M., Hanbury, A.: Patent retrieval. Found. Trends Inf. Retrieval **7**(1), 1–97 (2013)
9. Sowa, J.: Conceptual graphs. In: van Harmelen, F., Lifschitz, V., Porter, B. (eds.) Handbook of Knowledge Representation, pp. 213–237. Elsevier (2008)
10. Woods, W.: What's in a link: foundations for semantic networks. In: Bobrow, D., Collins, A. (eds.) Representation and Understanding, pp. 35–82. Academic Press, New York (1975)
11. Peirce, C.S.: Manuscripts on existential graphs. In: Collected Papers of Charles Sanders Peirce, vol. 4, pp. 320–410. Harvard University Press, Cambridge (1906)
12. Mineau, G., Missaoui, R., Godinx, R.: Conceptual modelling for data and knowledge management. Data Knowl. Eng. **33**(2), 137–168 (2000)
13. Croitoru, M., et al.: Conceptual graphs based information retrieval in HealthAgents. In: Twentieth IEEE International Symposium on Computer-Based Medical Systems, Maribor, Slovenia (2007)
14. Campbell, K.E., Musen, M.A.: Representation of clinical data using SNOMED III and conceptual graphs. In: Proceedings of the Annual Symposium Computer Applications in Medical Care (1992)

15. Bernauer, J.: Conceptual graphs as an operational model for descriptive findings. In: Proceedings of the Annual Symposium on Computer Application in Medical Care (1991)
16. Ounis, I., Pa©ca, M.: Modeling, indexing and retrieving images using conceptual graphs. In: Quirchmayr, G., Schweighofer, E., Bench-Capon, T.J.M. (eds.) DEXA 1998. LNCS, vol. 1460, pp. 226–239. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0054484
17. Mulhem, P., Leow, W.K., Lee, Y.K.: Fuzzy conceptual graphs for matching images of natural scenes. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence, Seattle, WA (2001)
18. Ryan, K., Mathews, B.: Matching conceptual graphs as an aid to requirements re-use. In: Proceedings of the IEEE International Symposium on Requirements Engineering, San Diego (1993)
19. Gerbé, O., Keller, R.K., Mineau, G.W.: Conceptual graphs for representing business processes in corporate memories. In: Mugnier, M.-L., Chein, M. (eds.) ICCS-ConceptStruct 1998. LNCS, vol. 1453, pp. 401–415. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0054931
20. Yang, S., Soo, V.W.: Extract conceptual graphs from plain text in patent claims. Eng. Appl. Artif. Intell. **25**(4), 874–887 (2012)
21. Zhong, J., Zhu, H., Li, J., Yu, Y.: Conceptual graph matching for semantic search. In: Priss, U., Corbett, D., Angelova, G. (eds.) ICCS-ConceptStruct 2002. LNCS, vol. 2393, pp. 92–106. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45483-7_8
22. Rao, P., Devi, S.L., Rosso, P.: Automatic identification of concepts and conceptual relations from patents using machine learning methods. In: Proceedings of the 10th International Conference on Natural Language Processing ICON-2013 (2013)
23. Rao, P., Devi, S.L.: Automatic identification of conceptual structures using deep Boltzmann machines. In: Proceedings of the 7th Forum for Information Retrieval Evaluation, FIRE 2015, Gandhinagar, India (2015)
24. Levesque, H., Brachman, R.: Knowledge Representation and Reasoning. Elsevier/Morgan Kaufmann, Amsterdam (2004)
25. Delugach, H.S.: Implementation and visualization of conceptual graphs in CharGer. Int. J. Conceptual Struct. Smart Appl. **2**(2), 1–19 (2014)
26. Gipp, B., Meuschke, N., Gernandt, A.: Decentralized trusted timestamping using the crypto currency bitcoin. In: Proceedings of the iConference 2015, Newport Beach, California (2015)