UDC 004.838.2
IRSTI 20.19.21

COMPARATIVE ANALYSIS OF KNOWLEDGE BASES: CONCEPTNET VS CYC

A. Y. Nuraliyeva[1], S. S. Daukishov[2], R.T. Nassyrova[3]
[1,2]Kazakh-British Technical University, Almaty, Kazakhstan
[3]IT Analyst, Philip Morris Kazakhstan, Almaty, Kazakhstan

[1]nuraliyevaassel@gmail.com, ORCID 0000-0001-6451-3743
[2]suleimen.daukishov@gmail.com, ORCID 0000-0002-3784-4456
[3]nassyrova.r@gmail.com, ORCID 0000-0002-7968-3195

Abstract. Data summarization, question answering, text categorization are some of the tasks knowledge bases are used for. A knowledge base (KB) is a computerized compilation of information about the world. They can be useful for complex tasks and problems in NLP and they comprise both entities and relations. We performed an analytical comparison of two knowledge bases which capture a wide variety of common-sense information - ConceptNet and Cyc.

Manually curated knowledge base Cyc has invested more than 1,000 man-years over the last two decades building a knowledge base that is meant to capture a wide range of common-sense skills. On the other hand, ConceptNet is a free multilingual knowledge base and crowdsourced knowledge initiative that uses a large number of links to connect commonplace items. When looking for common sense logic and answers to questions, ConceptNet is a great place to start.

In this research, two well-known knowledge bases were reviewed - ConceptNet and Cyc - their origin, differences, applications, benefits and disadvantages were covered. The authors hope this paper would be useful for researchers looking for more appropriate knowledge bases for word embedding, word sense disambiguation and natural-language communication.

Key words: knowledge base, natural language processing, data summarization, question answering, text categorization

Intoduction

A lot of research papers [1, 2, 3, 4] are devoted to the construction of Text to Knowledge computing systems. All such projects are united by directions of Deep Learning and technologies - Knowledge Base (KB) and Natural Language Processing (NLP). KB is a repository of information, where it is stored not separately, but in the context of other data units. By extracting data from natural language, we mean a machine that matches nouns with their entities and sentences with their statements. Such a transformation is possible not for any subject area but only where texts are subject to logical discourse and operate on facts (for example, jurisprudence, pharmaceuticals and other hard sciences).

Communication in natural languages is only possible if there is a considerable amount of general knowledge about the world that is shared between the various participants. By creating a knowledge base of general knowledge about the world, as well as specific knowledge in a particular field, they can be useful for a range of complex tasks and problems in NLP.

A knowledge base (KB) is a computerized compilation of information about the world that usually comprises both objects and information, facts about them. Data summarisation [5], named object disambiguation [6], question answering [7, 8], text categorization [9, 10, 11], coreference resolution [12, 13], plagiarism detection [14] are some of the tasks that KBs are used for.

In addition to text processing, structured vocabulary experience with wide coverage is useful for land applications such as geographic information systems and situated robots [15]. The majority of early methods to create KBs were manual. With the advancement of the Web, there are an increasing number of approaches to automatically generate KBs by collecting data from Web corpora. YAGO, DBpedia, Wikidata, NELL, and Google's Information Vault are some of the most well-known approaches. Some of these methods concentrate on Wikipedia, a free online encyclopedia.

Knowledge needed to understand any subject of interest is a key skill for interpreting and deciphering the ever-changing world in the information society. Since much knowledge is conveyed by linguistic communication, whether oral or written, understanding how words are used to convey meaning is critical. Lexical information can be found in a variety of formats, including unstructured terminologies, glossaries, machine-readable dictionaries (e.g. LDOCE [16]), and full computational vocabularies and ontologies (e.g. WordNet [17] and Cyc [18]). Manually developing certain services, on the other hand, is a daunting job. It takes decades to complete and must be done all over again with any new language. Non-English language resources frequently have much less coverage. As a result, analysis in resource-rich languages like English is clearly favored.

Well-known knowledge bases have been developed as a result of the creation of information retrieval approaches focused on the availability of a semantic framework for free text on the internet, such as Google Knowledge Graph [19], NELL [20], YAGO [21], and DBpedia [22]. These databases were created by major companies, such as Wikipedia websites [23], or on a large scale on the internet. Manually curated knowledge bases have also been created, such as the Cyc project [24], which has a small concept space of just 120,000 concepts. The Cyc knowledge base [25], according to Cycorp Inc., is the broadest, deepest, and most comprehensive repository ever built. The Cyc project has spent the last two decades – roughly 1,000 man-years – developing a knowledge base that is intended to capture a broad spectrum of common sense expertise. NELL [21], KnowItAll [26], and ReVerb [27] are examples of automatic approaches. They include many concepts found in scientific texts, but they fall far short of covering many concepts found in scientific articles.

In addition, ConceptNet is a free multilingual KB that uses a rich collection of links to bind everyday entities [28]. ConceptNet is a perfect place to start while searching for common sense reasoning and responses to questions. Finally, BabelNet, which incorporates definitions and relationships from WordNet, the world's biggest semantic lexicon, and Wikipedia, the most frequently cited collaborative encyclopedia.

Today, there are three important challenges in technologies of KB exposed in Cyc and ConceptNet, namely:

• Knowledge is logically contradictory and does not form a theory.

• The prediction for knowledge is poorly defined - the probabilistic assessment of the applicability of knowledge drops dramatically in the process of logical inference.

• Predictions derived from knowledge are statistically ambiguous.

The solution to these problems is discussed, for example, in the work of L. De Raedt and K. Kersting "Probabilistic logic learning". Their solution is closely related to the issue of probabilistic-logical learning, which is the integration of relational or logical representations, probabilistic inference and learning. The author [29] has shown that the solution of the above three problems is possible within the framework of a theory linking two approaches - probabilistic and logical.

Comparative analysis

ConceptNet

ConceptNet [30], a large-scale and publicly available knowledge base comprising millions of common sense statements presented in natural language [31]. ConceptNet was developed as part of the MIT Media Lab's Open Mind Common Sense initiative [32], a crowdsourced knowledge project that began in 1999. ConceptNet's graphical system of linked words/nodes for representing information is particularly useful for textual reasoning over natural language documents. The ConceptNet graph structure, which describes information in patterns of linked word/phrase nodes, is especially useful for textual reasoning over natural language documents. Figure 1 shows an overview of how such information is structured in ConceptNet.

Many practical word processing tasks for real documents are supported by ConceptNet without the need for additional statistical learning, such as thematic list (for example, a news article containing terms like "guns," "demand money," and "get away" may suggest the themes of "robbery" and "crime"), affect-sensitivity (for example, this letter is sad and angry), and analogy (e.g. "scissors", "razor", "nail clipper" and "sword" are perhaps similar to "knife" because they are all "sharp" and can be used to "cut something"),a short description, causal projection, cold grouping of records.

ConceptNet is an information graph that links natural language words and phrases to statements through marked edges. Its information is gathered and revised on a regular basis from a number of outlets, including expert tools, crowdsourcing, and purpose-built games. It is meant to present general language comprehension information in order to enhance natural language applications by helping them to better appreciate the meanings of the words people use.

ConceptNet provides vocabulary and world information from a variety of sources in a variety of languages, as well as acting as a link between knowledge tools. ConceptNet, for example, provides links to URLs defining astronomy in WordNet, Wiktionary, OpenCyc, and DBPedia in addition to its own knowledge of the English language. Embedded word construction is not the only use of ConceptNet, but it is a form of doing so that has strong advantages and is consistent with current distributive semantics study.

For decades, Cyc [19] has been constructing a predicate logic-based common sense information ontology. DBPedia [33] collects material from Wikipedia infoboxes, resulting in a vast number of statistics, mostly for named individuals of Wikipedia articles. Since its content is not publicly accessible, Google Knowledge Database [34] is probably the largest and most general knowledge network.

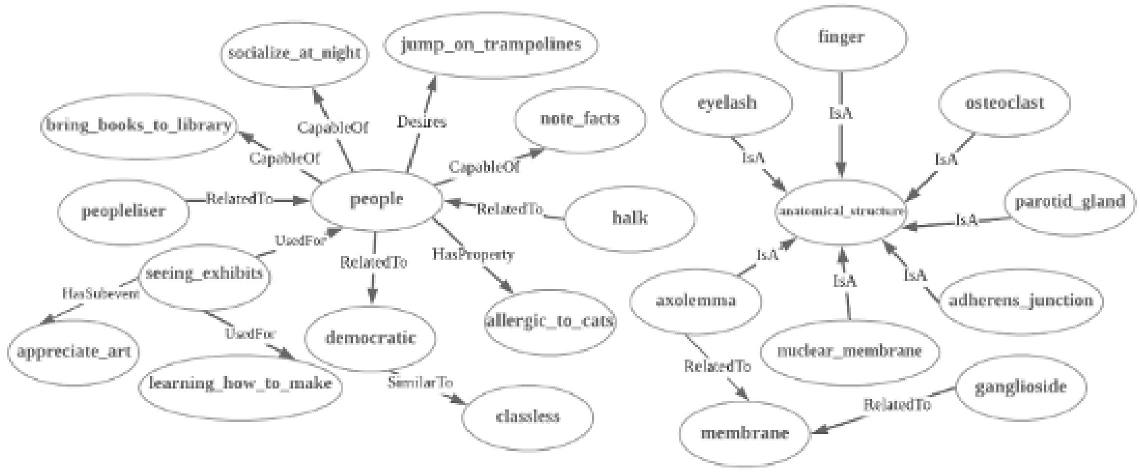

Figure 1 - Fragment of the ConceptNet Knowledge Graph

In contrast to these other tools, ConceptNet's function is to include a broad, free information graph that focuses on the common sense definitions of words as they are used in natural language. Because of the emphasis on words, it is particularly well suited to the concept of representing word meanings as vectors, which is a ConceptNet strength. Word combinations are represented as dense unit vectors of real numbers, with closely related vectors having semantic relationships. This representation is appealing because it portrays context as a continuous space in which relatedness and resemblance can be evaluated.

ConceptNet employs a closed class of relationships, such as IsA, UsedFor, and CapableOf, that are structured to represent relationships independent of the language or source of the words they connect. ConceptNet seeks to match its information services with the 36 partnerships that make up its core package. The intent of these generalised relations is close to that of WordNet relations like hyponym and meronym [35]. Although the edges in ConceptNet are directional, certain connections, such as SimilarTo, are designated as symmetric in ConceptNet 5.5, and the directionality of these edges is irrelevant. Since word embedding will learn from what ConceptNet understands, this knowledge base continues to play an important role in a field that has come to focus on word embedding. As shown by recent findings, ConceptNet may improve the robustness and correlation of word embedding with human judgment.

ConceptNet differs from other common information bases in a number of ways. ConceptNet, in contrast to WordNet [36], which focuses on maintaining lexicographic knowledge and the interaction between words and their context, retains a semantic network system that is intended to collect sensible sense sentences. ConceptNet, in fact, has more relationship forms than WordNet. Cyc [19], a knowledge base that focuses on standardizing common sense for effective logical reasoning, is another comparable knowledge base. ConceptNet, on the other hand, is optimized for inferring from natural language texts and, unlike Cyc, is not a proprietary system.

ConceptNet captures a wide variety of common sense information (as does Cyc), but it does so in a more user-friendly way than higher-order logical notation. ConceptNet has quickly become a useful dataset and resource for different forms of machine learning and NLP over the last decade, owing to these advantages [37, 38, 39].

NLP learning algorithms, especially those focused on embedding terms will benefit greatly from ConceptNet familiarity of graphical construction. ConceptNet can be used to create more effective semantic spaces than distributive semantics.

Various use cases with ConceptNet are given in numerous articles. For example, the paper "Semantic generation mechanism of news images based on Concept-Net" [40] suggests a news image description mechanism based on the ConceptNet knowledge graph. The model authors provide consists of two parts - extracting the image content and rendering with NLP to generate a description of the news image. Another one paper [41] describes enhanced story representation by ConceptNet to predict story endings. The authors propose to improve the representation of stories by simplifying the sentences to key concepts and then modeling the latent relationship between the key ideas within the story. Such enhanced sentence representation, when used with pre-trained language models, makes significant gains in prediction accuracy without using the biased validation data.

Besides, there are several papers about utilizing ConceptNet for sentiment analysis. For example, the paper "KGPChamps at SemEval-2019 Task 3: A deep learning approach to detect emotions in the dialog utterances" [42] describes an approach to solve a task where, given a textual dialogue, the emotion have to be classified according the utterance as one of the following emotion classes: Happy, Sad, Angry or Others. To solve this problem, authors experiment with ConceptNet and word embeddings generated from bi-directional long short-term memory (LSTM) taking input characters. As another example, Ramanathan, Vallikannu, and T. Meyyappan use sentiment analysis to get people feedback about Oman tourism by social media messages in their paper [43]. The authors created their own Oman tourism ontology based on ConceptNet. Entities are identified from the tweets using POS tagger and compared with concepts in the domain specific ontology. After that, the sentiment of the extracted entities are determined by the combined sentiment lexicon approach. Finally, semantic orientations of domain specific features are combined relating to the domain.

In addition, there are some use cases regarding ConceptNet in the educational area. One of them is described in the paper by Su, Ming-Hsiang, Chung-Hsien Wu, and Yi Chang [44] where authors propose an approach to generate follow-up questions based on a populated domain ontology in a conversational interview coaching system. The purpose was to produce the follow-up questions, which are related to the meaning based on the background knowledge in a populated domain ontology. Initially, a convolutional neural network was applied for selecting a key sentence from the user answer. Then the authors used the neural tensor network to model the relationship between the subjects and objects in the resource description framework triple in each predicate from the ConceptNet for domain ontology population. Another work [45] reports on a crowdsourcing experiment conducted with the help of the V-TREL vocabulary trainer (Telegram chatbot) to gather knowledge on word relations to expand ConceptNet. V-TREL offers vocabulary training exercises generated from ConceptNet and collects, assesses the learners' answers to extend ConceptNet with new words.

Moreover, ConceptNet was used in game development. For instance, one of the works introduces the Pokerator, a generator of creative Pokemon names and descriptions, based on user input [46]. The names are generated by mixing words based on syllables or characters according to a bigram language model. A concomitant description is generated by filling a template with ConceptNet answers. Another example is the paper by Lo, Chun Hei, and Luyang Lin, where the purpose is to refine 20Q, a computerized game of twenty yes-or-no questions that asks the player to think of something and the system tries to guess what they are thinking [47]. Authors provide possible research directions, mainly under the formulation of the problem as reinforcement learning. They also investigate methods and potential challenges of incorporating the use of KB for the game.

One more interesting use case is an android application named Talking Diary with a new approach for note classification and scheduling to enable mobile users to automatically organize their daily routine tasks with a single audio note [48]. The application contains three modules: auto audio note classification, auto audio note scheduling, and working hour's calculator. The proposed model of classifier computes similarity score by extracting N-gram weights from ConceptNet to execute classification.

Another research related to robotics proposes a new approach to connect household robot sensor data to Linked Data in order to give robotic agents semantic product information about objects that can be found in their environment, so that the action to be performed with a given object can be inferred [49]. For this, authors use the robot's belief state when recognizing a product and link it to a product ontology that follows Semantic Web standards. Then they use the product class information to fetch further information from ConceptNet that contains action information (e.g. laundry detergent is used for laundering). At last, the action results are mapped to internally known actions of the robotic agent, so that it knows which action can be done with the perceived object.

Cyc

The Cyc Knowledge Base (KB) [19] is the most comprehensive and deepest source of common sense knowledge ever created, with orders of magnitude more content than anyone else. Cyc uses real-world axioms to think about the world and explain the data, but KB is not a database. About 10,000 predicates, millions of sets and definitions, and over 25 million claims are used in Cyc's knowledge base. Cyc will easily prove trillions of bits of used information about the real world when combined with inference engines.

Douglas Lenat began the project at MCC (Microelectronics and Computer Technology Corporation) in July 1984, and the Cyc ontology quickly expanded to about 100,000 terms during the project's first decade, from 1984 to 1994, and included around 1,500,000 words as of 2017. This ontology included 416,000 collections (types, forms, natural species), slightly more than a million individuals comprising 42,500 predicates (relations, characteristics, areas, resources, functions), and roughly a million widely recognised organizations.

A significant number of additional concepts are indirectly used in the Cyc ontology. Cyc's KB of general rules and common sense concepts, which contains these ontological concepts, expanded from 1 million in 1994 to 24.5 million in 2017, which took over 1000 man-years to construct. It's necessary to note that Cyc's ontology engineers tend to hold these statistics as minimal as practicable rather than inflate them, as long as the information base's deductive closure is not impaired.

At different levels of generality, the information in the Cyc KB can be subdivided into loosely clustered, interrelated "blocks" of knowledge as shown in Figure 2. Cyc's interpretation of metaphysics is the broadest block of knowledge, grading down to very basic knowledge of loosely defined domains.Cyc is a long-term artificial intelligence (AI) project that seeks to build a systematic ontology and KB that encompasses fundamental principles and laws on how the world functions. Cyc focuses on tacit information that other AI systems can take for granted in the hopes of capturing common sense knowledge. This is in contrast to evidence that can be discovered on the internet, whether by a search engine or Wikipedia. When faced with new circumstances, Cyc encourages semantic reasoners to execute human reasoning and be less fragile.

The architecture of Cyc's inference engine distinguishes two problems: epistemological and heuristic, that is, what should be in the Cyc KB and how could Cyc efficiently extract arguments hundreds of steps deep in a sea of tens of millions of axioms. The CycL language and well-understood logical inferences could suffice for the former. Cyc used an agent group architecture in the second, in which specialized reasoning modules, each with their own data structure and algorithm, would "raise their hand" if they could work successfully on either of the open sub-problems. There were 20 heuristic level modules (HL) in 1994 [50]. Then there were over 1,050 HL modules in 2017.

The Cyc project's main aim has been to build a large knowledge base containing a stock of formalized context knowledge useful for a range of logic and problem solving activities in various domains since the beginning. Although systems possessing only specialized information of particular domains have produced remarkable outcomes, Cyc's thesis is that these systems are unstable [51] and impossible to apply to modern or unknown problems or problem domains [52]. This is particularly true in areas concerning natural-language communication or responses to questions [53], where the problem domain's breadth is often difficult or impossible to completely describe in advance.

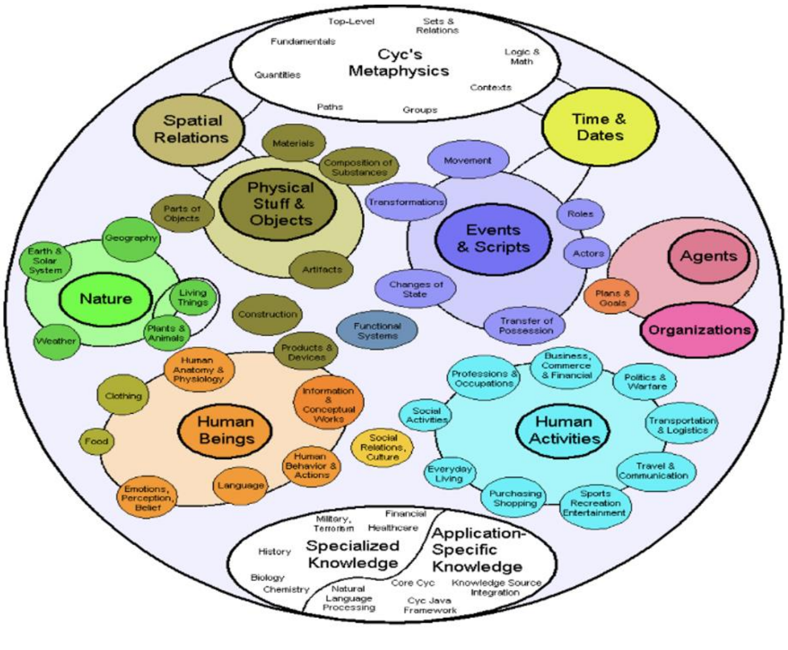

Figure 2 - Cyc knowledge base topic map

This knowledge base is intended to help potential knowledge representation and thinking activities that are unexpected (and also unanticipated). At the time, ontologies or slices of ontology from the Cyc KB were used in a variety of applications, including CycSecure [54], the very early Thesaurus Manager, and the Cyc "Digital Aristotle" program [51].

Applications of the Cyc scheme include:

1. Terrorist identification based on information

The Terrorism Identification Knowledge Base [55] is a Cyc program or framework currently in progress with the goal of preserving all information about "terrorist" organizations, their participants, and those who operate their groups, as well as ideology, founders, guarantors, alliances, programs, locations, capitals, competencies, priorities, events, tactics, and full metaphors for specific terrorist acts. The data is saved as mathematical reasoning files that lead to machine sympathies and cognitive functions.

2. Encyclopedia

An encyclopedia is being created that overlays Cyc keywords on the Wikipedia sheets that are occupied from the sides.

3. Clinical studies: methanalysis

By extending Cyc's ontology and KB about 2%, Cycorp and Cleveland Clinic Foundation (CCF) have built a system to answer clinical researchers' ad hoc queries [56]. The system employs a series of CycL (higher-order logic) pieces of open variables, which are then constrained by medical domain awareness, human-like common sense, grammar, and other factors. There is a way to merge these fragments into a single formal question that is semantically expressive.

4. Healthcare decision support system

The study [57] proposes a knowledge-based system using ontological engineering (by adopting the Cyc method) to assist the creation of a robust foundation for establishing a decision-making support system for the proper diagnosis and management of diseases (e.g. Typhoid Fever, Malaria, Diarrhoea Diseases, Pneumonia, Anaemia) in Sunyani Municipality.

5. Healthcare: effective personalized cancer treatments

The Big C ('C' for Cyc) is a system designed to (semi-)automatically obtain, integrate, and use complex mechanism models related to cancer biology by means of automated reading and a hyper-detailed refinement process resting on Cyc's logical representations and powerful inference mechanisms [58]. Authors' goal is to assist cancer research and treatment with the scale and attention to detail that only computer implementations can provide.

6. Network Security

CycSecureTM is a network risk assessment and network monitoring application that relies on knowledge-based artificial intelligence technologies to improve on traditional network vulnerability assessment [54]. CycSecure integrates public reports of software faults from online databases, data collected automatically from computers on a network and hand-ontologized information about computers and computer networks. This information is stored in Cyc KB and processed by the Cyc inference engine and planner to deliver detailed analyses of the security and vulnerability of networks.

7. Assist in semantic summarization

An abstract description of record data is generated which focuses on the Cyc platform for development. The system's knowledge base and inference engine help it to abstract new ideas that aren't mentioned directly in the document. It makes use of the text's semantic characteristics and syntactic structure. Furthermore, the knowledge base offers subject-matter insight and helps the system to manipulate relationships between ideas in records, which is highly advantageous.

8. Smart AI education

The study [59] offers sixth-grade mathematics learning-by-teaching (LBT) system BELLA built by slightly extending the ontology and KB of Cyc. The "teachable agent" Elle begins with an understanding of the domain content close to the human student's. There is a super-agent (Cyc) which knows the domain content well. BELLA builds up a mental model of the human student by observing them interact with Elle to decide what Elle's current mental model should be to help the user to overcome their current confusions.

Cyc's advantages is that it uses common sense intelligence to dismiss conclusions such as a man becoming pregnant or a 20-year-old having served for 22 years if the inputs suggest so.

To overcome syntactic ambiguities, one must first comprehend the essence of sentences. This enables machines to interpret and comprehend documents written in natural human languages without being caught up in elliptical expressions, overt inconsistencies, or vague sentences.

The Cyc project has been criticized by many AI researchers for lacking a theoretical foundation. However, Lenat contends that while the inference engine can perform deductive reasoning based on database information, it cannot perform induction, that is, it cannot take new data and produce new ideas or relationships. Furthermore, it is 'crystalline' in the sense that assertions are not probabilistically tested. Lenat replied to some of the objections by emphasizing that Cyc is not attempting to create a full artificial general intelligence (AGI) and that it can be quickly integrated into other AI initiatives.

Conclusion

Freebase, WordNet, DBpedia, and Yago are examples of KBs that have been successfully developed and applied to a variety of NLP problems. Machine learning and representational algorithms based on these KBs have advanced year after year in such diverse fields as information graph 'embedding,' question answering, and remote surveillance. ConceptNet and Cyc are the only two cases of common sense logic that we are aware of.

ConceptNet has been described as a semantic network and Cyc has been described as a resource that enables a better comprehension of sentence semantics in a variety of ways. Cyc works on creating a universal schema higher-order logic for expressing sensible semantic statements and can also be used to help reasoning structures [60, 61] that can make more complex logical inferences. New statements are applied to the KB on a regular basis using a mixture of automatic and manual methods.

For decades, Cyc has been constructing a predicate logic-based common sense information ontology. In contrast, ConceptNet's function is to include a broad, free information graph that focuses on the common sense definitions of words as they are used in natural language. Because of the emphasis on words, it is particularly well suited to the concept of representing word meanings as vectors, which is a ConceptNet strength.

Cyc, a knowledge base that focuses on standardizing common sense for effective logical reasoning, is another comparable knowledge base. ConceptNet, on the other hand, is optimized for inferring from natural language texts and, unlike Cyc, is not a proprietary system.

ConceptNet captures a wide variety of common sense information as does Cyc, but it does so in a more user-friendly way than higher-order logical notation. Cyc is created by experts, while ConceptNet is a crowdsourced knowledge project.

NLP learning algorithms, especially those focused on embedding terms, such as word representations in vector space [62], will benefit greatly from ConceptNet familiarity of graphical construction. ConceptNet can be used to create more effective semantic spaces than distributive semantics. On the other hand, Cyc's advantage is that it uses common sense intelligence to dismiss conclusions with no logic.

Regarding known use cases of ConceptNet and Cyc we can conclude that ConceptNet is better represented in the scientific community as there are more relevant articles describing use cases, than we found for Cyc. Besides, ConceptNet is more oriented towards the mass market because there are many use cases in online media, sentiment analysis, game development, and education. Use cases of Cyc are more narrow-profile, such as healthcare, clinical studies, and network security.

References

[1] http://www.opencyc.org [Access date 08042021]

[2] http://www.link.cs.cmu.edu/link/ [Access date 08042021]

[3] http://gate.ac.uk/ [Access date 08042021]

[4] https://github.com/tensorflow/models/tree/master/syntaxnet [Access date 08042021]

[5] Nastase V. Topic-driven multi-document summarization with encyclopedic knowledge and activation spreading. Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. Waikiki, Honolulu, Hawaii. (2008). 763–772.

[6] Bunescu R., Pasca M. Using encyclopedic knowledge for named entity disambiguation. Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics. Trento, Italy. (2006). 9–16.

[7] Harabagiu S.M., Moldovan D., Pasca M., Mihalcea R., Surdeanu M., Bunescu R., Girju

R., Rus V., Morarescu P. “Boosting knowledge for answer engines”. Proceedings of the Ninth Text REtrieval Conference. Gaithersburg, Maryland. (2000). 479–488.

[8] Lita L.V., Hunt W.A., Nyberg E. “Resource analysis for question answering”. Companion Volume to the Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics. Barcelona, Spain. (2004). 162–165.

[9] Gabrilovich E., Markovitch S. Wikipedia-based semantic interpretation for natural language processing. Journal of Artificial Intelligence Research (JAIR). 2009. 443–498.

[10] Wang P., Domeniconi C. Building semantic kernels for text classification using Wikipedia. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, Nevada. 2008. 713–721.

[11] Navigli R., Faralli S., Soroa A., de LacalleO.L., Agirre E. Two birds with one stone: Learning semantic models for Text Categorization and Word Sense Disambiguation. Proceedings of the Twentieth ACM Conference on Information and Knowledge Management. Glasgow, Scotland. (2011). 2317–2320.

[12] Ponzetto S.P., Strube M. Knowledge derived from Wikipedia for computing semantic relatedness. Journal of Artificial Intelligence Research. 2007. 181–212.

[13] Rahman A., Ng V. Narrowing the modeling gap: A cluster-ranking approach to coreference resolution. Journal of Artificial Intelligence Research. 2011. 469–521.

[14] Barron-Cedeno A., Rosso P., Agirre E., Labaka G. Plagiarism detection across distant language pairs. Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, China. (2010). 37–45.

[15] Bateman J. A., Hois J., Ross R., Tenbrink T. A linguistic ontology of space for natural language processing. Artificial Intelligence. 2010. 1027–1071.

[16] Proctor P. Longman Dictionary of Contemporary English. Longman Group, Harlow, UK. 1978.

[17] Miller G.A., Beckwith R., Fellbaum C.D., Gross D., Miller K. WordNet: an online lexical database. International Journal of Lexicography. 1990. 235–244.

[18] Lenat D.B. Cyc: A large-scale investment in knowledge infrastructure. Communications of the ACM. 1995.

[19] Agirre E., Soroa A. Personalizing PageRank for Word Sense Disambiguation. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Athens, Greece. (2009). 33–41.

[20] Von Ahn L., Kedia M., Blum M. Verbosity: A game for collecting common-sense facts. Proceedings of the Conference on Human Factors in Computing Systems. Canada. (2006). 75–78.

[21] Atserias J., Villarejo L., Rigau G., Agirre E., Carroll J., Magnini B., Vossen P. The MEANING multilingual central repository. Proceedings of the 2nd International Global WordNet Conference. Brno, Czech Republic. (2004). 80–210.

[22] Banerjee S., Pedersen T. Extended gloss overlap as a measure of semantic relatedness. Proceedings of the 18th International Joint Conference on Artificial Intelligence. Acapulco, Mexico. (2003). 805–810.

[23] https://www.wikipedia.org [Access date 10042021]

[24] Banko M., Cafarella M.J., Soderland S., Broadhead M., Etzioni O. Open information extraction from the Web. Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India. (2007). 2670–2676.

[25] Chan Y.S., Ng H.T., Zhong Z. NUS-ML: Exploiting parallel texts for word sense disambiguation in the English all-words tasks. Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007). Prague, Czech Republic. (2007). 253–256.

[26] Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. DBpedia – A crystallization point for the web of data. Journal of Web. Semanti. (2009). 154–165.

[27] Black W., Rodriguez S.E.H., Alkhalifa M., Vossen P., Pease A. Introducing the Arabic WordNet project. Proceedings of the 3rd International Global WordNet Conference. Jeju Island, South Kore. (2006). 295–299.

[28] Harabagiu S.M., Miller G.A., Moldovan D.I. WordNet 2 – a morphologically and semantically enhanced resource. Proceedings of the SIGLEX99 Workshop on Standardizing Lexical Resources. (1999). 1–8.

[29] Vityaev E.E. Knowledge extraction from data. Computer Knowledge Model of cognitive process. Novosibirsk 2006, 293 p.

[30] Hirschberg J., Manning C.D. Advances in natural language processing. Science 2015. 349(6245). 261–266.

[31] Speer R., Havasi C. Conceptnet 5: A large semantic network for relational knowledge. The people's web meets NLP, theory and applications of natural language processing. 2013. 161–176.

[32] Havasi C., Speer R., Arnold K., Lieberman H., Alonso J., Moeller J. Open mind common sense: Crowdsourcing for common sense. Proceedings of the 2nd AAAI Conference on Collaboratively Built Knowledge Sources and Artificial Intelligence. AAAI Press. (2010). 53.

[33] Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P., Hellmann S., Morsey M., Van Kleef P., Auer S., Bizer C. Dbpedia a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web Journal. 2014.

[34] GuiLin QI, GAO W.T. Huan. The research advances of knowledge graph. Technology Intelligence Engineering. 2017. 4–25.

[35] Pustejovsky J. The generative lexicon. Computational linguistics. 1991. 17(4). 409–441.

[36] Miller G.A. Wordnet: A lexical database for English. 1995. 38(11). 39–41.

[37] Shi B., Weninger T. Open-world knowledge graph completion. 2017.

[38] Chen M., Tian Y., Chang K.-W., Skiena S., Zaniolo C. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. 2018.

[39] Lin B.Y., Chen X., Chen J., Ren X. Kagnet: Knowledge-aware graph networks for commonsense reasoning. 2019.

[40] Chen, Zhongmin, Hao Xu, Shuo Gao, and Yongsheng Wang. Semantic generation mechanism of news images based on Concept-Net. In 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). IEEE (2021). 245-248.

[41] Huang, Shanshan, Kenny Q. Zhu, Qianzi Liao, Libin Shen, and Yinggong Zhao. Enhanced Story Representation by ConceptNet for Predicting Story Endings. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, (2020). 3277-3280.

[42] Patro, Jasabanta, Nitin Choudhary, Kalpit Chittora, and Animesh Mukherjee. KGPChamps at SemEval-2019 Task 3: A deep learning approach to detect emotions in the dialog utterances. In Proceedings of the 13th International Workshop on Semantic Evaluation, (2019). 241-246.

[43] Ramanathan, Vallikannu, and T. Meyyappan. Twitter text mining for sentiment analysis on people's feedback about oman tourism. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), IEEE (2019). 1-5.

[44] Su, Ming-Hsiang, Chung-Hsien Wu, and Yi Chang. Follow-Up Question Generation Using Neural Tensor Network-Based Domain Ontology Population in an Interview Coaching System. In INTERSPEECH. 2019. 4185-4189.

[45] Rodosthenous, Christos, Verena Lyding, Federico Sangati, Alexander König, Umair ul Hassan, Lionel Nicolas, Jolita Horbacauskiene, Anisia Katinskaia, and Lavinia Aparaschivei. Using crowdsourced exercises for vocabulary training to expand conceptnet. In Proceedings of The 12th Language Resources and Evaluation Conference, (2020). 307-316.

[46] Geissler, Dominique, Elisa Nguyen, Daphne Theodorakopoulos, and Lorenzo Gatti. Pokérator-Unveil your inner Pokémon. In 11th International Conference on Computational Creativity. (2020).

[47] Lo, Chun Hei, and Luyang Lin. Reinforcement Learning in 20Q Game with Generic Knowledge Bases.

[48] Munir, A. Hani, Abubakar Manzoor, and Utba Aziz. Talking Diary: A Novel Approach

for Automatic Audio Note Categorization and Event Scheduling for Android Application. 2020.

[49] Kümpel, Michaela, Anna de Groot, Ilaria Tiddi, and Michael Beetz. Using linked data to help robots understand product-related actions. In JOWO. 2020.

[50] Lenat D.B., Guha R. V. Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[51] Friedland N.S, Allen, P.G., Witbrock M., Angele J., Staab S., Israel D., Chaudhri,V., Porter B., Barker K., Clark P. Towards a Quantitative, Platform-Independent Analysis of Knowledge Systems, in Proc. of the 9th International Conference on the Principles of Knowledge Representation and Reasoning. Whistler, (2004). 507-515.

[52] Lenat, D. "Cyc: A Large-Scale Investment in Knowledge Infrastructure, CACM 38. 1995. 11.

[53] Hovy E., Hermjakob U., Lin C-Y., Ravichandran D. Using knowledge to facilitate factoid answer pinpointing Full text.. In Proc. of the 19th International Conference on Computational Linguistics. Taipei, Taiwan, (2002). 1.

[54] Shepard B., Matuszek C., Fraser C.B., Wechtenhiser W., Crabbe D., Güngördü Z., Jantos J., Hughes T., Lefkowitz L., Witbrock M., Lenat D., Larson E. A Knowledge-Based Approach to Network Security: Applying Cyc in the Domain of Network Risk Assessment. In Proc. of the 17th Innovative Applications of Artificial Intelligence Conference. Pittsburgh, PA. (2005).

[55] Deaton, Chris, Blake Shepard, Charles Klein, Corrinne Mayans, Brett Summers, Antoine Brusseau, Michael Witbrock, and Doug Lenat. The comprehensive terrorism knowledge base in cyc. In Proceedings of the 2005 International Conference on Intelligence Analysis. (2005).

[56] Lenat, Douglas, Michael Witbrock, David Baxter, Eugene Blackstone, Chris Deaton, Dave Schneider, Jerry Scott, and Blake Shepard. Harnessing Cyc to answer clinical researchers' ad hoc queries. AI Magazine 31, (2010). 3. 13-32.

[57] Appiah S., A. F. Adekoya C. Bapuuroh and C. Akowua-Kwakye. A Knowledge-Base of Prevalent Diseases in Sunyani Municipality, Ethiopia Using Ontological Engineering. J Health Med Informat 2019. 10(338). 2.

[58] Witbrock, Michael, Karen Pittman, Jessica Moszkowicz, Andrew Beck, Dave Schneider, and Douglas Lenat. Cyc and the big c: Reading that produces and uses hypotheses about complex molecular biology mechanisms. In Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence. (2015).

[59] Lenat, Douglas B., and Paula J. Durlach. Reinforcing math knowledge by immersing students in a simulated learning-by-teaching experience. International Journal of Artificial Intelligence in Education. 2014. 24(3). 216-250.

[60] Panton K., Matuszek C., Lenat D., Schneider D., Witbrock M., Siegel N., Shepard B. Common Sense Reasoning – From Cyc to Intelligent Assistant. 2006. 3864. 1–31.

[61] Ramachandran D., Reagan P., Goolsbey K. First-orderized researchcyc: Expressivity and efficiency in a common-sense ontology. 2005.

[62] Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. 2013.