Knowledge-based Artificial Intelligence

by Mike Bergman - Monday, November 17, 2014

http://www.mkbergman.com/1816/knowledge-based-artificial-intelligence/

What Goes Around, Comes Around, Only Now with Real Knowledge

A recent interview with a noted researcher, IEEE Fellow Michael I. Jordan, Pehong Chen Distinguished Professor at the University of California, Berkeley, provided a downplayed view of recent AI hype. Jordan was particularly critical of AI metaphors to real brain function and took the air out of the balloon about algorithm advances, pointing out that most current methods have roots that are decades long [1]. In fact, the roots of knowledge-based artificial intelligence (KBAI), the subject of this article, also extend back decades.

Yet the real point, only briefly touched upon by Jordan in his lauding of Amazon's recommendation service, is that the dynamo in recent AI progress has come from advances in the knowledge and statistical bases driving these algorithms. The improved digital knowledge bases behind KBAI have been the power behind these advances.

Knowledge bases are finally being effectively combined with AI, a dynamic synergy that is only now being recognized, let alone leveraged. As this realization increases, many forms of useful information structure in the wild will begin to be mapped to these knowledge bases, which will further extend the benefits we are are now seeing from KBAI.

Knowledge-based artificial intelligence, or KBAI, is the use of large statistical or knowledge bases to inform feature selection for machine-based learning algorithms used in AI. The use of knowledge bases to train the features of AI algorithms improves the accuracy, recall and precision of these methods. This improvement leads to perceptibly better results to information queries, including pattern recognition. Further, in a virtuous circle, KBAI techniques can also be applied to identify additional possible facts within the knowledge bases themselves, improving them further still for KBAI purposes.

It is thus, in my view, the combination of KB + AI that has led to the notable AI breakthroughs of the

past, say, decade. It is in this combination that we gain the seeds for sowing AI benefits in other areas, from tagging and disambiguation to the complete integration of text with conventional data systems. And, oh, by the way, the structure of all of these systems can be made inherently multi-lingual, meaning that context and interpretation across languages can be brought to our understanding of concepts.

Structured Dynamics is working to democratize a vision of KBAI that brings its benefits to any enterprise, using the same approaches that the behemoths of the industry have used to innovate knowledge-based artificial intelligence in the first place. How and where the benefits of such KBAI may apply is the subject of this article.

A Brief History of Knowledge-based Systems

Knowledge-based artificial intelligence is not a new idea. Its roots extend back perhaps to one of the first AI applications, Dendral. In 1965, nearly a half century ago, Edward Feigenbaum initiated Dendral, which became a ten-year effort to develop software to deduce the molecular structure of organic compounds using scientific instrument data. Dendral was the first expert system and used mass spectra or other experimental data together with a knowledge base of chemistry to produce a set of possible chemical structures. This set the outline for what came to be known as knowledge-based systems, which are one or more computer programs that reason and use knowledge bases to solve complex problems.

Indeed, it was in the area of expert systems that AI first came to the attention of most enterprises. According to Wikipedia,

Expert systems were designed to solve complex problems by reasoning about knowledge, represented primarily as if–then rules rather than through conventional procedural code. The first expert systems were created in the 1970s and then proliferated in the 1980s. Expert systems were among the first truly successful forms of AI software.

Expert systems spawned the idea of knowledge engineers, whose role was to interview and codify the logic of the chosen experts. But, expert systems proved to be expensive to build and difficult to maintain and tune. As the influence of expert systems waned, another branch emerged, that of knowledge-based engineering and their support for CAD- and CASE-type systems. Still, overall penetration to date of most knowledge-based systems can most charitably be described as disappointing.

The specific identification of "KBAI" was (to my knowledge) first made in a Carnegie-Mellon University report to DARPA in 1975 [2]. The source knowledge bases were broadly construed, including listings of hypotheses. The first known patent citing knowledge-based artificial intelligence is from 1992 [3]. Within the next ten years there were dedicated graduate-level course offerings on KBAI at many universities, including at least Indiana University, SUNY Buffalo, and Georgia Tech.

In 2007, Bosse? *et al.* devoted a chapter to KBAI in their book on information fusion, but still, at that time, the references were more generic [4]. However, by 2013, the situation was changing fast, as this quote from Hovy *et al.* indicates [5]:

"Recently, however, this stalemate [the so-called 'knowledge acquisition bottleneck] has begun to loosen

up. The availability of large amounts of wide-coverage semantic knowledge, and the ability to extract it using powerful statistical methods, are enabling significant advances in applications requiring deep understanding capabilities, such as information retrieval and question-answering engines. Thus, although the well-known problems of high cost and scalability discouraged the development of knowledge-based approaches in the past, more recently the increasing availability of online collaborative knowledge resources has made it possible to tackle the knowledge acquisition bottleneck by means of massive collaboration within large online communities. This, in turn, has made these online resources one of the main driving forces behind the renaissance of knowledge-rich approaches in AI and NLP – namely, approaches that exploit large amounts of machine-readable knowledge to perform tasks requiring human intelligence." (citations removed)

The waxing and waning of knowledge-based systems and its evolution over fifty years have led to a pretty well-defined space, even if not all component areas have achieved their commercial potential. Besides areas already mentioned, knowledge-based systems also include:

- Knowledge models — formalisms for knowledge representation and reasoning, and
- Reasoning systems — software that generates conclusions from available knowledge using logical techniques such as deduction and induction.

We can organize these subdomains as follows. Note particularly that the branch of KBAI (knowledge-based artificial intelligence) has two main denizens: recognized knowledge bases, such as Wikipedia, and statistical corpora. The former are familiar and evident around us; the latter are largely proprietary and not (generally) publicly accessible:
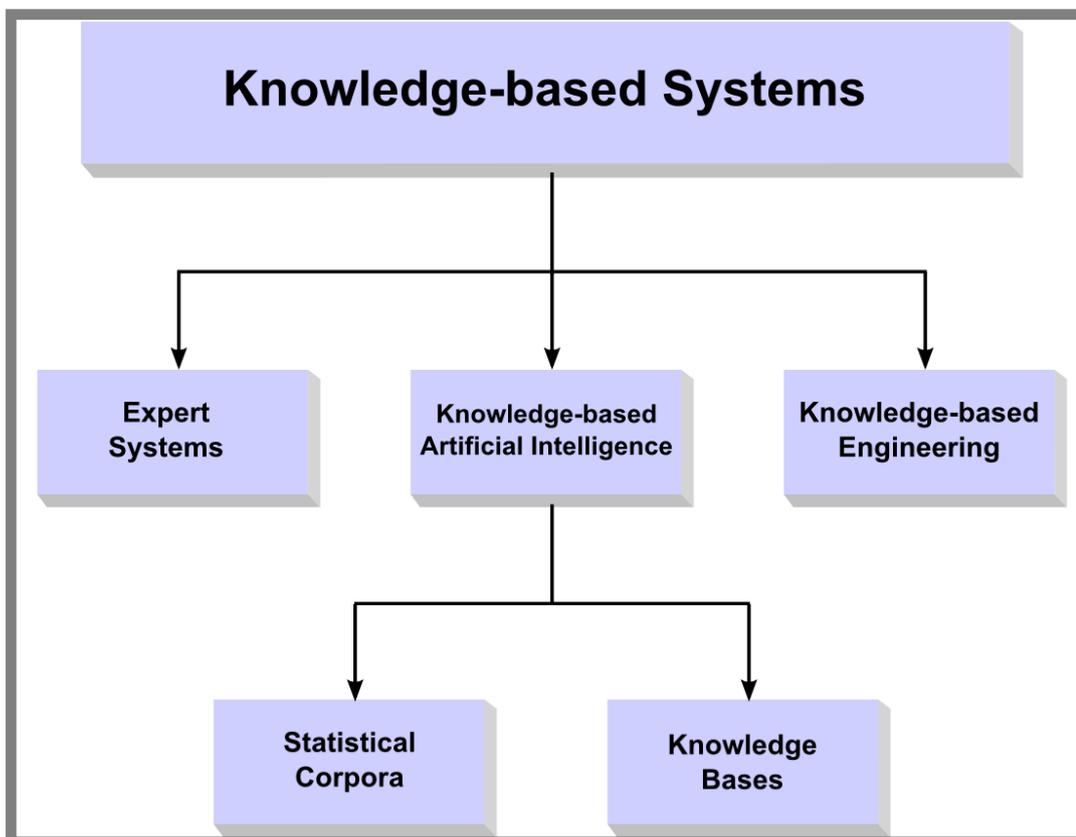
Some prominent

knowledge bases and statistical corpora are identified below. Knowledge bases are coherently organized information with instance data for the concepts and relationships covered by the domain at hand, all accessible in some manner electronically. Knowledge bases can extend from the nearly global, such as Wikipedia, to very specific topic-oriented ones, such as restaurant reviews or animal guides. Some electronic knowledge bases are designed explicitly to support digital consumption, in which case they are fairly structured with defined schema and standard data formats and, increasingly, APIs. Others may be electronically accessible and highly relevant, but the data is not staged in a easily-consumable way, thereby requiring extraction and processing prior to use.

The use and role of statistical corpora is harder to discern. Statistical corpora are organized statistical relationships or rankings that facilitate the processing of (mostly) textual information. Uses can range from entity extraction to machine language translation. Extremely large sources, such as search engine indexes or massive crawls of the Web, are most often the sources for these knowledge sets. But, most are applied internally by those Web properties that control this big data.

The Web is the reason these sources — both statistical corpora and knowledge bases — have proliferated, so the major means of consuming them is via Web services with the information defined and linked to URIs.

My major thesis has been that it is the availability of electronically accessible knowledge bases, exemplified and stimulated by Wikipedia [6], that has been the telling factor in recent artificial intelligence advances. For example, there are at least 500 different papers that cite using Wikipedia for various natural language processing, artificial intelligence, or knowledge base purposes [7]. These papers began to stream into conferences about 2005 to 2006, and have not abated since. In turn, the various techniques innovated for extracting more and more structure and information from Wikipedia are being applied to other semi-structured knowledge bases, resulting in a true renaissance of knowledge-based processing for AI purposes. These knowledge bases are emerging as the information substrate under many recent computational advances.

Knowledge Bases in Relation to Overall Artificial Intelligence

A few months ago I pulled together a bit of an interaction diagram to show the relationships between major branches of artificial intelligence and structures arising from big data, knowledge bases, and other organizational schema for information:

What we are seeing is a system emerging whereby multiple portions of this diagram interact to produce innovations. Take, for example, Apple's Siri [8], or Google's Google Now or the many similar systems that have emerged on smartphones. Spoken instructions are decoded to text, which is then parsed and evaluated for intent and meaning and then posed to a general knowledge base. The text results are then modulated back to speech with the answer in the smartphone's speakers. The pattern recognition at the front and back end of this workflow has been made better though statistical datasets derived from phonemes and text. The text understanding is processed via natural language processing and semantic technologies [16], with the question understanding and answer formulation coming from one or more knowledge bases.

This remarkable chain of processing is now almost taken for granted, though its commercial use is less than five years old. For different purposes with different workflows we see effective question answering and diagnosis with systems like IBM's Watson [9] and structured search results from Google's Knowledge Graph [10]. Try posing some questions to Wolfram Alpha and then stand back and be impressed with the data visualization. Behind the scenes, pattern recognition from faces to general images or thumbprints further is eroding the distinction between man and machine. Google Translate now covers language translation between 60 human languages [11] — and pretty damn effectively, too. All major Web players are active in these areas, from Amazon's recommendation system [12] to Facebook [42], Microsoft [13], Twitter [14] or Baidu [15].

Though not universal, most all recent AI advances leveraging knowledge bases have utilized Wikipedia in one way or another. Even Freebase, the core of Google's Knowledge Graph, did not really blossom as a separate data crowdsourcing concern until its former owner, Metaweb, decided to bring Wikipedia into its system. Many other knowledge bases, as noted below, are also derivatives or enhancements to Wikipedia in one way or another.

I believe the reasons for Wikipedia's influence have arisen from its nearly global scope, its mix of semi-structured data and text, its nearly 200 language versions, and its completely open and accessible nature. Regardless, it is also certainly true that techniques honed with Wikipedia are now being applied to a diversity of knowledge bases. We are also seeing an appreciation start to grow in how knowledge bases can enhance the overall AI effort.

Useful Statistical and Knowledge Sources

The diagram on knowledge-based systems above shows two kinds of databases contributing to KBAI: statistical corpora or databases and true knowledge bases. The statistical corpora tend to be hidden behind proprietary curtains, and also more limited in role and usefulness than general knowledge bases.

Statistical Corpora

The statistical corpora or databases tend to be of a very specific nature. While lists of text corpora and many other things may contribute to this category, the ones actually in commercial use tend to be quite focused in scope, very large, and designed for bespoke functionality. A good example, and one that has been contributed for public use, is the Web 1T 5-gram data set [17]. This data set, contributed by Google for public use in 2006, contains English word n-grams and their observed frequency counts. N-grams

capture word tokens that often coincide with one another, from single words to phrases. The length of the n-grams ranges from unigrams (single words) to five-grams. The database was generated from approximately 1 trillion word tokens of text from publicly accessible Web pages.

Another example of statistical corpora are what is used in Google's Translate capabilities [11]. According to Franz Josef Och, who was the lead manager at Google for its translation activities and an articulate spokesperson for statistical machine translation, a solid base for developing a usable language translation system for a new pair of languages should consist of a bilingual text corpus of more than a million words, plus two monolingual corpora each of more than a billion words. Statistical frequencies of word associations form the basis of these reference sets. Google originally seeded its first language translators with multiple language texts from the United Nations [18].

Such lookup or frequency tables in fact can shade into what may be termed a knowledge base as they gain more structure. NELL, for example (and see below), contains a relatively flat listing of assertions extracted from the Web for various entities; it goes beyond frequency counts or relatedness, but does not have the full structure of a general knowledge base like Wikipedia [19]. We thus can see that statistical corpora and knowledge bases in fact reside on a continuum of structure, with no bright line to demark the two categories.

Nonetheless, most statistical corpora will never be seen publicly. Building them requires large amounts of input information. And, once built, they can offer significant commercial value to their developers to drive various machine learning systems and for general lookup.

Knowledge Bases

There are literally hundreds of knowledge bases useful to artificial intelligence, most of a restricted domain nature. Listed below, partially informed by Suchanek and Weikum's work [20], are some of the broadest ones available. Note that many leverage or are derivatives of or extensions to Wikipedia:

- BabelNet — is a multilingual lexicalized semantic network and ontology automatically created by linking Wikipedia to WordNet [21]
- Biperpedia — is an ontology with 1.6M (class, attribute) pairs and 67K distinct attribute names, a totally unique resource, but one that is not publicly available [22]
- ConceptNet — is a semantic network with concepts as nodes and edges that are assertions of common sense about these concepts [23]
- Cyc — is an artificial intelligence project that attempts to assemble a comprehensive ontology and knowledge base of everyday common sense knowledge, with the goal of enabling AI applications to perform human-like reasoning [24]
- DBpedia — extracts structured content from the information created as part of the Wikipedia, principally from its infoboxes [25]
- DeepDive — employs statistical learning and inference to combine diverse data resources and best-of-breed algorithms in order to construct knowledge bases from hundreds of millions of Web pages [26]
- EntityCube — is a knowledge base built from the statistical extraction of structured entities, named entities, entity facts and relations from the Web [27]
- Freebase — is a large collaborative knowledge base consisting of metadata composed mainly by its

community members, but centered initially on Wikipedia; Freebase is a key input component to Google's Knowledge Graph [28]

- GeoNames — is a geographical database that contains over 10,000,000 geographical names corresponding to over 7,500,000 unique features [29]
- ImageNet — is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images [30]
- KnowItAll — is a variety of domain-independent systems that extract information from the Web in an autonomous, scalable manner [31]
- Knowledge Vault (Google) — is a Web-scale probabilistic knowledge base that combines extractions from Web content (obtained via analysis of text, tabular data, page structure, and human annotations) with prior knowledge derived from existing knowledge repositories; it is not publicly available [32]
- LEVAN (Learning About Anything) — is a fully-automated approach for learning extensive models for a wide range of variations (*e.g.*, actions, interactions, attributes and beyond) from images for any concept, leveraging the vast resources of online books [43]
- NELL (Never-Ending Language Learning system) — is a semantic machine learning system developed at Carnegie Mellon University that identifies a basic set of fundamental semantic relationships between a few hundred predefined categories of data, such as cities, companies, emotions and sports teams [19]
- Probase — is a universal, probabilistic taxonomy that contains 2.7 million concepts harvested from a corpus of 1.68 billion web pages [33]
- UMBEL — is an upper ontology of about 28,000 reference concepts and a vocabulary for aiding that ontology mapping, including expressions of likelihood relationships [34]
- Wikidata — is a common source of certain data types (for example, birth dates) which can be used by Wikimedia projects such as Wikipedia [35]
- WikiNet — is a multilingual extension of the facts found in the multiple language versions of Wikipedia [36]
- WikiTaxonomy — is a large-scale taxonomy derived from the category relationships in Wikipedia [37]
- Wolfram Alpha — is a computational knowledge engine made available by subscription as an online service that answers factual queries directly by computing the answer from externally sourced "curated data"
- WordNet — is a lexical database for the English language that groups words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members [38]
- YAGO — is extracted from Wikipedia (*e.g.*, categories, redirects, infoboxes), WordNet (*e.g.*, synsets, hyponymy), and GeoNames [39].

What Work is Being Done and the Future

It is instructive to inspect what kinds of work or knowledge these bases are contributing to the AI enterprise. The most important contribution, in my mind, is structure. This structure can relate to the subsumption (is-a) or part of (mereology) relationships between concepts. This form of structure contributes most to understanding the taxonomy or schema of a domain; that is, its scaffolding of concepts. This structure helps orient the instance data and other external structures, generally through

some form of mapping.

The next rung of contribution from these knowledge bases is in the nature of the relations between concepts and their instances. These form the predicates or nature of the relationships between things. This kind of contribution is also closely related to the attributes of the concepts and the properties of the things that populate the structure. This kind of information tends to be the kind of characteristics that one sees in a data record: a specific thing and the values for the fields by which it is described.

Another contribution from knowledge bases comes from identity and disamgibuation. Identity works in that we can point to authoritative references (with associated Web identifiers) for all of the individual things and properties in our relevant domain. We can use these identities to decide the "canonical" form, which also gives us a common reference for referring to the same things across information sources or data sets. We also gain the means for capturing the various ways that anything can be described, that is the synonyms, jargon, slang, acronyms or insults that might be associated with something. That understanding helps us identify the core item at hand. When we extend these ideas to the concepts or types that populate our relevant domain, we can also begin to establish context and other relationships to individual things. When we encounter the person "John Smith" we can use co-occurring concepts to help distinguish John Smith the plumber from John Smith the politician or John Smith the policeman. As more definition and structure is added, our ability to discriminate and disambiguate goes up.

Some of this work resides in the interface between schema (concepts) and the instances (individuals) that populate that schema, what I elsewhere have described as the work between the T-Box and A-Box of knowledge bases [40]. In any case, with richer understandings of how we describe and discern things, we can now begin to do new work, not possible when these understandings were lacking. We can now, for example, do semantic search where we can relate multiple expressions for the same things or infer relationships or facets that either allow us to find more relevant items or better narrow our search interests.

With true knowledge bases and logical approaches for working with them and their structure, we can begin doing direct question answering. With more structure and more relationships, we can also do so in rather sophisticated ways, such as identifying items with multiple shared characteristics or within certain ranges or combinations of attributes.

Structured information and the means to query it now gives us a powerful, virtuous circle whereby our knowledge bases can drive the feature selection of AI algorithms, while those very same algorithms can help find still more features and structure in our knowledge bases. The interaction between AI and the KBs means we can add still further structure and refinement to the knowledge bases, which then makes them still better sources of features for informing the AI algorithms:
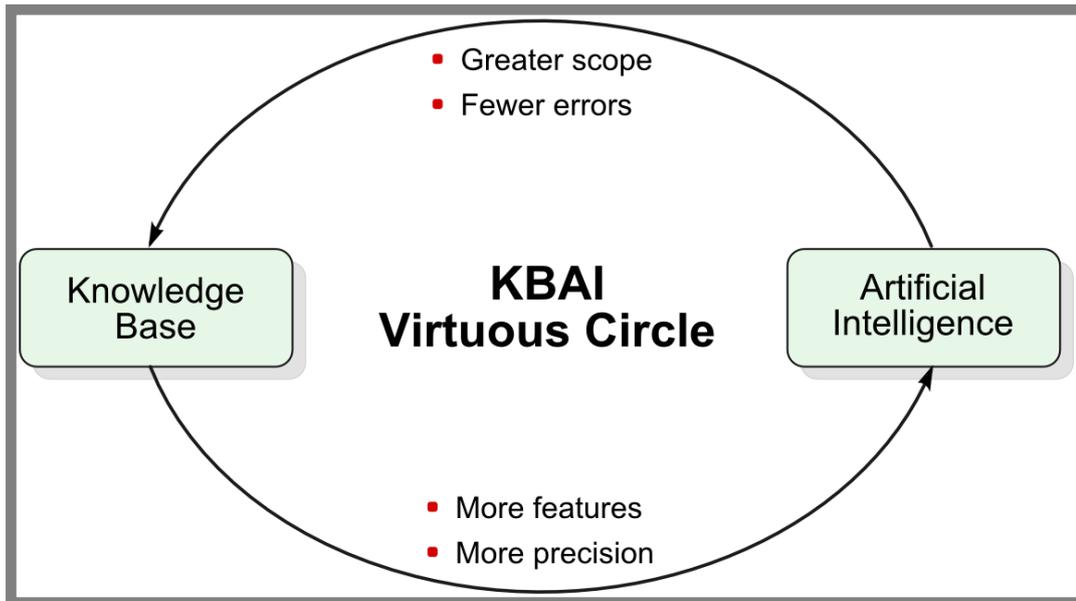
Once this threshold of feature generation is reached, we now have a virtuous dynamo for knowledge discovery and management. We can use our AI techniques to refine and improve our knowledge bases, which then makes it easier to improve our AI algorithms and incorporate still further external information. Effectively utilized KBAI thus becomes a generator of new information and structure.

This virtuous circle has not yet been widely applied beyond the early phases of, say, adding more facts to Wikipedia, as some of our examples above show. But these same basic techniques can be applied to the very infrastructural foundations of KBAI systems in such areas as data integration, mapping to new external structure and information, hypothesis testing, diagnostics and predictions, and the myriad of other uses to which AI has been hoped to contribute for decades. The virtuous circle between knowledge bases and AIs does not require us to make leaps and bounds improvements in our core AI algorithms. Rather, we need only stoke our existing AI engines with more structure and knowledge fuel in order to keep the engine chugging.

The vision of a growing nexus of KBAI should also prove that efficiencies and benefits also increase through a power function of the network effect, similar to what I earlier described in the Viking algorithm [41]. We know how we can extract further structure and benefit from Wikipedia. We can see how such a seed catalyst can also be the means for mapping and relating more specific domain knowledge bases and structure. The beauty of this vision is that we already can see the threshold benefits from a decade of KBAI development. Each new effort — and there are many — will only act to add to these benefits, with each new increment contributing more than the increment that came before. That sounds to me like productivity, and a true basis for wealth creation.

---

[1] See Lee Gomes, 2014. "Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts," in *IEEE Spectrum*, 20 Oct 2014. See also http://www.kdnuggets.com/2014/11/berkeley-michael-jordan-big-data-transformative-not-delusion.html.

[2] Lee D. Erman and Victor R. Lesser, 1975. A Multi-level Organization for Problem Solving Using Many Diverse, Cooperating Sources of Knowledge, DARPA Report AD-AO12-919, Carnegie-Mellon University, Pittsburgh, Pa. 15213 , March, 1975, 24 pp. See http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA012916. The authors followed this up with a 1978

book, *System engineering techniques for artificial intelligence systems*. Academic Press, 1978.

[3] See "Method for converting a programmable logic controller hardware configuration and corresponding control program for use on a first programmable logic controller to use on a second programmable logic controller," US Patent No 5142469 A, August 25, 1992.

[4] E?loi Bosse?, Jean Roy, Steve Wark, Eds., 2007. *Concepts, Models, and Tools for Information Fusion*, Artech House Publishers, 2007-02-28, SKU-13/ISBN: 9781596939349. Chapter 11 is devoted to "Knowledge-Based and Artificial Intelligence Systems."

[5] Eduard Hovy, Roberto Navigli, and Simone Paolo Ponzetto, 2013. "Collaboratively Built Semi-Structured Content and Artificial Intelligence: The Story So Far," *Artificial Intelligence* **194** (2013): 2-27. See http://wwwusers.di.uniroma1.it/~navigli/pubs/AIJ_2012_Hovy_Navigli_Ponzetto.pdf.

[6] See M.K. Bergman, "SWEETpedia," listing of Wikipedia research articles, on *AI3:::Adaptive Information* blog, January 25, 2010. The listing as of its last update included 246 articles.

[7] See the combined references between [6] and [20]; also, see Wikipedia's own "Wikipedia in Academic Studies."

[8] Wade Roush, 2010. "The Story of Siri, from Birth at SRI to Acquisition by Apple–Virtual Personal Assistants Go Mobile." *Xconomy. com* **14** (2010).

[9] Special Issue on "This is Watson", 2012. I*BM Journal of Research and Development* **56(3/4)**, May/June 2012. See also this intro to Watson video.

[10] Amit Singhal, 2012. " Introducing the Knowledge Graph: Things, not Strings," *Google Blog*, May 16, 2012

[11] Thomas Schulz, 2013. "Google's Quest to End the Language Barrier," September 13, 2013, Spiegel Online International. See http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html. Also see, Alon Halevy, Peter Norvig, and Fernando Pereira, 2009. "The Unreasonable Effectiveness of Data," in *IEEE Intelligent Systems*, March/April 2009, pp 8-12.

[12] Greg Linden, Brent Smith, and Jeremy York, 2003. "Amazon. com Recommendations: Item-to-item Collaborative Filtering," *Internet Computing, IEEE* 7, no. **1** (2003): 76-80.

[13] Athima Chansanchai, 2014. "Microsoft Research Shows off Advances in Artificial Intelligence with Project Adam," on *Next at Microsoft* blog, July 14, 2014. Also, see [33].

[14] Jimmy Lin and Alek Kolcz, 2012. "Large-Scale Machine Learning at Twitter," *SIGMOD*, May 20–24, 2012, Scottsdale, Arizona, US. See http://www.dcs.bbk.ac.uk/~DELL/teaching/cc/paper/sigmod12/p793-lin.pdf.

[15] Robert Hof, 2014. "Interview: Inside Google Brain Founder Andrew Ng's Plans To Transform Baidu," *Forbes Online*, August 28, 2014.

[16] Alok Prasad and Lee Feigenbaum, 2014, "How Semantic Web Tech Can Make Big Data Smarter," in *CMSwire*, Oct 6, 2014.

[17] The Web 1T 5-gram data set is available from the Linguistic Data Corporation, University of Pennsylvania.

[18] Franz Josef Och, 2005. "Statistical Machine Translation: Foundations and Recent Advances," *The Tenth Machine Translation Summit*, Phuket, Thailand, September 12, 2005.

[19] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr, and Tom M. Mitchell, 2010. "Toward an Architecture for Never-Ending Language Learning," in *AAAI*, vol. **5**, p. 3. 2010.

[20] Fabian M. Suchanek and Gerhard Weikum, 2014. "Knowledge Bases in the Age of Big Data Analytics," *Proceedings of the VLDB Endowment* 7, no. **13** (2014).

[21] Roberto Navigli and Simone Paolo Ponzetto, 2012. "BabelNet: The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network," Artificial Intelligence **193** (2012): 217-250.

[22] Rahul Gupta, Alon Halevy, Xuezhi Wang, Steven Euijong Whang, and Fei Wu, 2014. "Biperpedia: An Ontology for Search Applications," . *Proceedings of the VLDB Endowment* **7**(7), 2014.

[23] Robert Speer and Catherine Havasi, "Representing General Relational Knowledge in ConceptNet 5," *LREC* 2012, pp. 3679-3686.

[24] Douglas B. Lenat and R. V. Guha, 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Addison-Wesley, 1990 ISBN 0-201-51752-3.

[25] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives 2007. "DBpedia: A Nucleus for a Web of Open Data," presented at *ISWC 2007*, Springer Berlin Heidelberg, 2007.

[26] Feng Niu, Ce Zhang, Christopher Ré, and Jude W. Shavlik, 2012. "DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference," in *VLDS*, pp. 25-28. 2012.

[27] Zaiqing Nie, J-R. Wen, and Wei-Ying Ma, 2012. " Statistical Entity Extraction From the Web," in *Proceedings of the IEEE* 100, no. **9** (2012): 2675-2687.

[28] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor, 2008. "Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 1247-1250. ACM, 2008.

[29] Mark Wick and Bernard Vatant, 2012. "The GeoNames Geographical Database," available from the *World Wide Web* at http://geonames.org (2012).

[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, 2009. " ImageNet: A Large-scale Hierarchical Image Database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248-255.

[31] Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates, 2005. "Unsupervised Named-entity Extraction from the Web: An Experimental Study," *Artificial Intelligence* **165**, no. 1 (2005): 91-134.

[32] Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun and Wei Zhang, 2014. "Knowledge Vault: a Web-scale Approach to Probabilistic Knowledge Fusion," KDD 2014.

[33] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q. Zhu, 2012. "Probase: A Probabilistic Taxonomy for Text Understanding," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 481-492. ACM, 2012.

[34] M.K. Bergman, 2008. "UMBEL: A Subject Concepts Reference Layer for the Web", *Slideshare*, July 2008.

[35] Denny Vrande?i? and Markus Krötzsch, 2014. "Wikidata: A Free Collaborative Knowledgebase," *Communications of the ACM* 57, no. **10** (2014): 78-85.

[36] Vivi Nastase and Michael Strube. 2013. "Transforming Wikipedia into a Large Scale Multilingual Concept Network," *Artificial Intelligence* **194** (2013): 62-85.

[37] Simone Paolo Ponzetto, and Michael Strube, 2007. "Deriving a Large Scale Taxonomy from Wikipedia," in *AAAI*, vol. **7**, pp. 1440-1445. 2007.

[38] Christiane Fellbaum and George Miller, eds., 1998. *WordNet: An Electronic Lexical Database*, MIT Press, 1998.

[39] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum, 2013. "YAGO2: a Spatially and Temporally Enhanced Knowledge Base from Wikipedia," *Artificial Intelligence* **194** (2013): 28-61; and Fabien M. Suchanek, Gjergji Kasneci, Gerhard Weikum, 2007. "YAGO: a Core of Semantic Knowledge," in *Proceedings of the 16th international Conference on World Wide Web*, pp. 697-706. ACM, 2007.

[40] See the T-Box and A-Box discussion with particular reference to artificial intelligence on M.K. Bergman, 2014. "Big Structure and Data Interoperability," on *AI3:::Adaptive Information* blog, August 18, 2014.

[41] M.K. Bergman, 2014. "The Value of Connecting Things – Part II: The Viking Algorithm," on *AI3:::Adaptive Information* blog, September 14, 2014.

[42] Cade Metz, 2013. "Facebook Taps 'Deep Learninig' Giant for New AI Lab," on *Wired.com*, December 9, 2013.

[43] Santosh Divvala, Ali Farhadi and Carlos Guestrin, 2014. "Learning Everything about Anything: Webly-Supervised Visual Concept Learning," in *CVPR 2014*.

PDF generated by *AI3:::Adaptive Information* blog