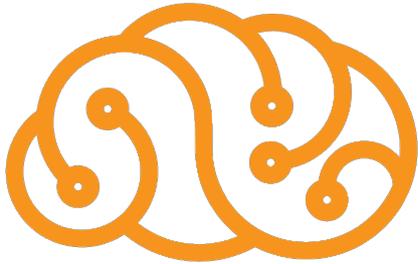


Cognonto is on the Hunt for Big AI Game

by Mike Bergman - Tuesday, September 20, 2016

<http://www.mkbergman.com/1981/cognonto-is-on-the-hunt-for-big-ai-game/>



cognonto

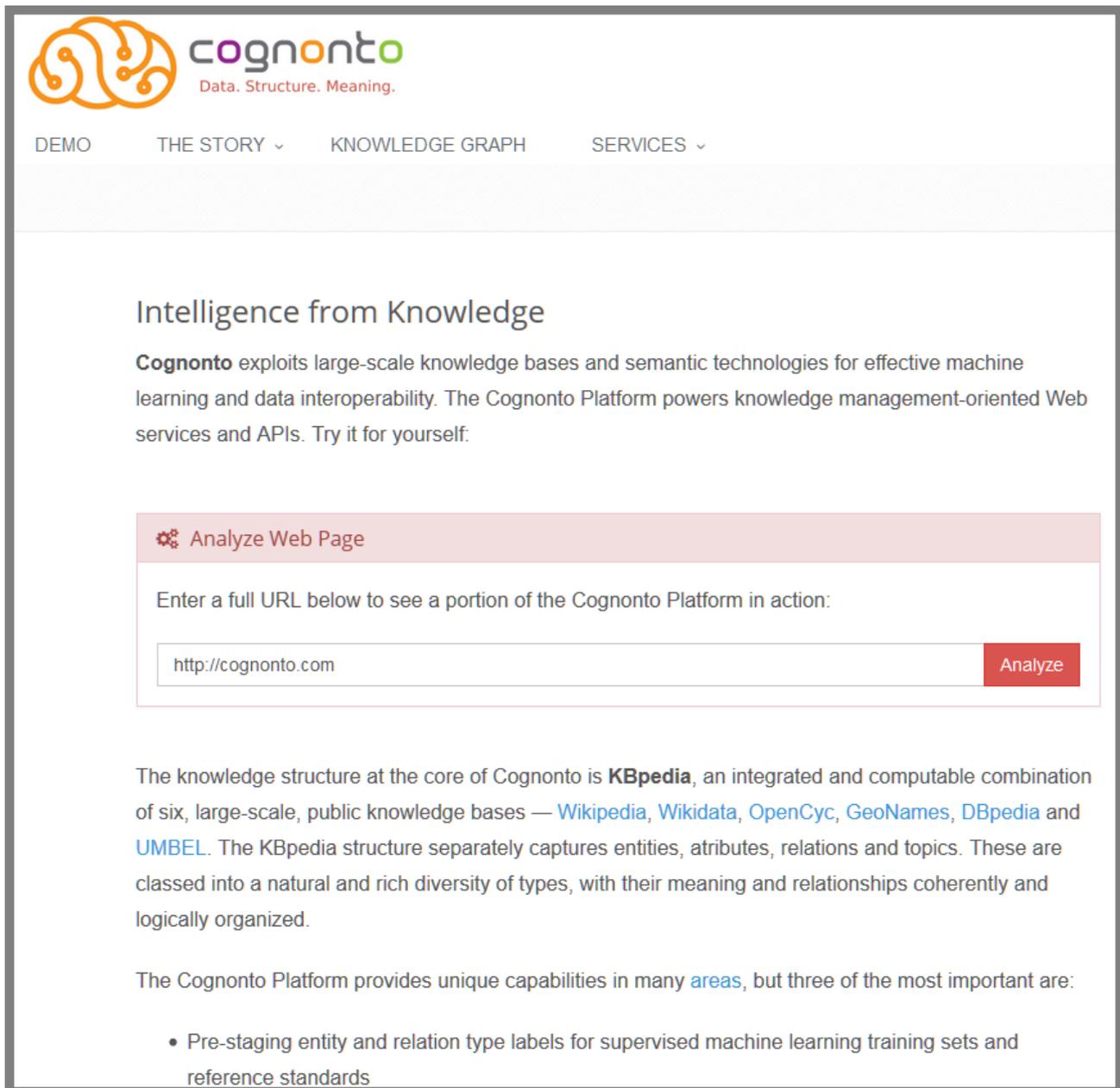
Six Large-scale Knowledge Bases Interact to Help Automate Machine Learning Setup

[Fred Giasson](#) and I [today announced](#) the unveiling of a new venture, [Cognonto](#). We have been working on this venture very hard for at least the past two years. But, frankly, Cognonto represents bringing into focus ideas and latent opportunities that we have been seeing for much, much longer.

The fundamental vision for Cognonto is to organize the information in large-scale knowledge bases so as to efficiently support knowledge-based artificial intelligence ([KBAI](#)), a topic I have been [writing about much](#) over the past year. Once such a vision is articulated, the threads necessary to bring it to fruition come into view quickly. First, of course, the maximum amount of information possible in the source knowledge bases needs to be made digital and represented with semantic Web technologies such as RDF and OWL. Second, since no source alone is adequate, the contributing knowledge bases need to be connected and made to work with one another in a logical and consistent manner. And, third, an overall schema needs to be put in place that is coherent and geared specifically to knowledge representation and machine learning.

The result from achieving these aims is to greatly lower the time and cost to prepare inputs to, and improve the accuracy in, machine learning. This result applies particularly to supervised machine learning for knowledge-related applications. But, if achieved, the resulting rich structure and extensive features also lend themselves to unsupervised and deep learning, as well as to provide a powerful substrate for schema mapping and data interoperability.

Today, we've now made sufficient progress on this vision to enable us to release Cognonto, and the KBpedia knowledge structure at its core. Combined with local data and schema, there is much we can do with the system. But another exciting part is that the sky is the limit in terms of honing the structure, growing it, and layering more AI applications upon it. Today, with Cognonto's release, we begin that process.



Screen Shot of the Entry Point for the Cognonto Demo

You can begin to see the power and the structure yourself via Cognonto's [online demo](#), as shown above, which showcases a portion of the system's functionality.

Problem and Opportunity

Artificial intelligence (AI) and machine learning are revolutionizing knowledge systems. Improved algorithms and faster graphics chips have been contributors. But the most important factor in knowledge-based AI's renaissance, in our opinion, has been the availability of massive digital datasets for the training of machine learners.

Wikipedia and data from search engines are central to recent breakthroughs. Wikipedia is at the heart of Siri, Cortana, the former Freebase, DBpedia, Google's Knowledge Graph and IBM's Watson, to name just a prominent few AI question answering systems. Natural language understanding is showing impressive gains across a range of applications. To date, all of these examples have been the result of bespoke efforts. It is very expensive for standard enterprises to leverage these knowledge resources on their own.

Today's practices pose significant upfront and testing effort. Much latent knowledge remains unexpressed and not easily available to learners; it must be exposed, cleaned and vetted. Further upfront effort needs to be spent on selecting the features (variables) used and then to accurately label the positive and negative training sets. Without "gold standards" — at still more cost — it is difficult to tune and refine the learners. The cost to develop tailored extractors, taggers, categorizers, and natural language processors is simply too high.

So recent breakthroughs demonstrate the promise; now it is time to systematize the process and lower the costs. The insight behind Cognonto is that existing knowledge bases can be staged to automate much of the tedium and reduce the costs now required to set up and train machine learners for knowledge purposes. Cognonto's mission is to make knowledge-based artificial intelligence (KBAI) cheaper, repeatable, and applicable to enterprise needs.

[Cognonto](#) (a *portmanteau* of 'cognition' and 'ontology') exploits large-scale knowledge bases and semantic technologies for machine learning, data interoperability and mapping, and fact and entity extraction and tagging. Cognonto puts its insight into practice through a knowledge structure, KBpedia, designed to support AI, and a management framework, the Cognonto Platform, for integrating external data to gain the advantage of KBpedia's structure. We automate away much of the tedium and reduce costs in many [areas](#), but three of the most important are:

- Pre-staging labels for entity and relation types, essential for supervised machine learning training sets and reference standards; KBpedia's structure-rich design is further useful for unsupervised and deep learning;
- Fine-grained entity and relation type taggers and extractors; and
- Mapping to external schema to enable integration and interoperability of structured, semi-structured and unstructured data (that is, everything from text to databases).

The KBpedia Knowledge Structure

KBpedia is a computable knowledge structure resulting from the combined mapping of six, large-scale, public knowledge bases — [Wikipedia](#), [Wikidata](#), [OpenCyc](#), [GeoNames](#), [DBpedia](#) and [UMBEL](#). The KBpedia structure separately captures entities, attributes, relations and topics. These are classed into a natural and rich diversity of types, with their meaning and relationships logically and coherently organized. This diagram, one example from the [online demo](#), shows the topics captured for the main Cognonto page in relation to the major typologies within KBpedia:

Example Network Graph as One Part of the Demo Results

Each of the six knowledge bases has been mapped and re-expressed into the KBpedia Knowledge Ontology. KKO follows the universal categories and logic of the 19th century American mathematician and philosopher, [Charles Sanders Peirce](#), the subject of [my last article](#). KKO is a computable knowledge graph that supports inference, reasoning, aggregations, restrictions, intersections, and other logical operations. KKO's logic basis provides a powerful way to represent individual things, classes of things, and how those things may combine or emerge as new knowledge. You can inspect the [upper portions of the KKO structure](#) on the Cognonto Web site. Better still, if you have an ontology editor, you can download and inspect the [open source KKO](#) directly.

KBpedia contains nearly 40,000 *reference concepts* (RCs) and about 20 million entities. The combination of these and KBpedia's structure results in nearly 7 billion logical connections across the system, as these KBpedia statistics (current as of today's version 1.02 release) show:

Measure	Value
No KBpedia reference concepts (RCs)	38,930
No. mapped vocabularies	27
Core knowledge bases	6
Extended vocabularies	21
No. mapped classes	138,868
Core knowledge bases	137,203
Extended vocabularies	1,665
No. typologies (SuperTypes)	63
Core entity types	33
Other core types	5
Extended	25
Typology assignments	545,377
No. aspects	80
Direct entity assignments	88,869,780
Inferred entity aspects	222,455,858
No. unique entities	19,643,718
Inferred no of entity mappings	2,772,703,619

Total no. of "triples"	3,689,849,726
Total no. of inferred and direct assertions	6,482,197,063

First Release KBpedia Statistics

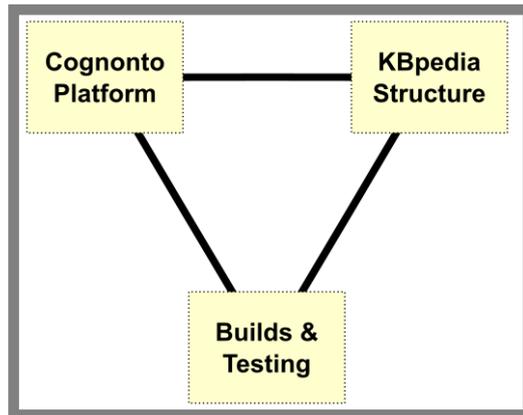
About 85% of the RCs are themselves entity types — that is, 33,000 natural classes of similar entities such as ‘astronauts’ or ‘breakfast cereals’ — which are organized into about [30 “core” typologies](#) that are mostly disjoint (non-overlapping) with one another. KBpedia has extended mappings to a further [20 other vocabularies](#), including schema.org, Dublin Core, and others; client vocabularies are typical additions. The typologies provide a flexible means for slicing-and-dicing the knowledge structure; the entity types provide the tie-in points to KBpedia’s millions of individual instances (and for your own records). KBpedia is expressed in the semantic Web languages of OWL and RDF. Thus, most W3C standards may be applied against the KBpedia structure, including for linked data, a standard option.

KBpedia is purposefully designed to enable meaningful splits across any of its structural dimensions — concepts, entities, relations, attributes, or events. Any of these splits — or other portions of KBpedia's [rich structure](#) — may be the computable basis for training taggers, extractors or classifiers. Standard NLP and machine learning reference standards and statistics are applied during the parameter-tuning and learning phases. Multiple learners and recognizers may also be combined as different signals to an ensemble approach to overall scoring. Alternatively, KBpedia's slicing-and-dicing capabilities may drive export routines to use local or third-party ML services under your own control.

Though usable in a standalone mode, only slices of KBpedia may be applicable to a given problem or domain, which then most often need to be extended with local data and schema. Cognonto has services to incorporate your own domain and business data, critical to fulfill domain purposes and to respond to your specific needs. We transform your external and domain data into KBpedia's canonical forms for interacting with the overall structure. Such data may include other public databases, but also internal, customer, product, partner, industry, or research information. Data may range from unstructured text in documents to semi-structured tags or metadata to spreadsheets or fully structured databases. The formats of the data may span hundreds of document types to all flavors of spreadsheets and databases.

Platform and Technology

Cognonto's modular technology is based on Web-oriented architectures. All functionality is exposed via Web services and programmatically in a microservice design. The technology for Cognonto resides in three inter-related areas:



- **Cognonto Platform** - the technology for storing, accessing, mapping, visualizing, querying, managing, analyzing, tagging, reporting and machine learning using KBpedia;
- **KBpedia Structure** – the central knowledge structure of organized and mapped knowledge bases and their millions of instances; and
- **Build Infrastructure** - repeatable and modifiable build and coherence and consistency testing scripts, including reference standards.

The Cognonto Web services may be manipulated directly from the command line or via cURL calls, or by simple HTML interfaces, by SPARQL, or programmatically. The Web services are written in Clojure and follow literate programming practices.

The base KBpedia [knowledge graph](#) may be [explored interactively](#) across billions of combinations with sample exports of its content. Here is the example for automobile.

version 1.02 See a problem? Tell us. Reference Concept Search

car

<http://kbpedia.org/kko/rc/Automobile>

motorcar | automobile | auto | cars | autos | motorcars | automobiles

A specialization of [motor vehicle](#). Each instance of [car](#) is a four-wheeled road vehicle designed for carrying about two to eight passengers. Notable specializations of [car](#) include [sports car](#), [pickup truck](#) and [Van](#). This collection does not include motorcycles or tractor-trailers (since both motorcycles and tractor-trailers are not four-wheeled vehicles), nor does it include buses (since buses are designed to carry more passengers than instances of [car](#) are designed to carry).

Core Structure

These are the structural linkages among the core knowledge bases for this RC: ... [more](#)

Sub-Classes

Direct	Inferred
wikidata:Q741490	5
wikidata:Q850270	56
wikidata:Q3231690	5.4K

Super-Classes

Direct	Inferred
opencyc:Mx4r1jUXeq00EdmAAAACs0uFOQ	8.7M 0.4M
opencyc:Mx4r2krpS0xcEduAADggVbxzQ	5.2M 0.1M
opencyc:Mx4r4Yj7IjVuQdeUnb1quULZig	8.7M 0.4M

Example Screen Shot for a Portion of the Knowledge Graph Results

There is a lot going on with many results panels and with links throughout the structure. There is a ['How to'? for the knowledge graph](#) if you really want to get your hands dirty.

These platform, technology, and knowledge structure capabilities combine to enable us to offer services across the full spectrum of KBAI applications, including:

- [Machine Learning Services](#);
- [Taggers, Extractors, Classifiers](#);
- [Content Harvesters/Publishers](#);
- [Data Prep and Staging](#);
- [Schema Mapping and Integration](#);
- [Knowledge Graph Services](#);
- [Dedicated SaaS](#); and
- [Platform Deployment](#).

Cognonto is a foundation for doing serious knowledge-based artificial intelligence.

Today and Tomorrow

Despite the years we have been working on this, it very much feels like we are at the beginning. There is so much more that can be done.

First, we need to continue to wring out errors and mis-assignments in the structure. We estimate an accuracy error rate of 1-2% currently, but that still represents millions of potential errors. The objective is not to be more accurate than alternatives, which we already are, but to be the most effective foundation possible for training machine learners. Further cleaning will result in still better standards and mappings. Throughout the [interactive knowledge graph](#) we have a button for submitting errors; please so submit if you see any problems!

Second, we are seeing the value of exposing structure, and the need to keep doing so. Each iteration of structure gets easier, because prior ones may be applied to automate much of the testing and vetting effort for the subsequent ones. Structure provides the raw feature (variable) grist used by machine learners. We have a very long punch list of where we can effectively add more structure to KBpedia.

And, last, we need to extend the mappings to more knowledge bases, more vocabularies, and more schema. This kind of integration is really what smooths the way to data integration and interoperability. Virtually every problem and circumstance requires including local and external information.

We know there are many important uses — and an upside of potential — for codifying knowledge bases for AI and machine learning purposes. [Drop me a line](#) if you'd like to discuss how we can help you leverage

your own domain and business data using knowledge-based AI.

PDF generated by *AI3::Adaptive Information* blog