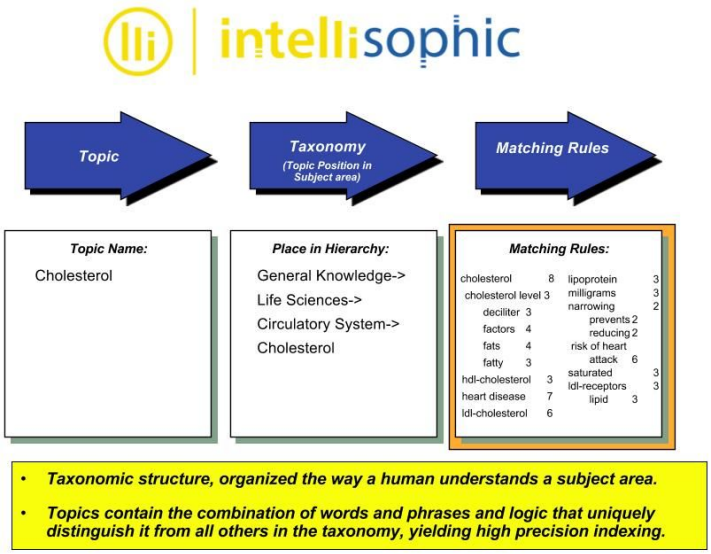
Semantic Data for
Intelligent Agents and Assistants

"The Semantic Web is not a new computing environment but rather an extension of the existing Web. Semantic data that provides a machine-readable "meaning" of information is layered over the information that is provided for people. The power of semantic data is that software can process it." Peter Sweeny, **The History of the Semantic Web is the Future of Intelligent Assistants.**

Intellisophic provides taxonomic content and technology that enables the systematic use of unstructured information. Our software and API platforms provide the foundation for extracting machine understandable data from conversations and text. This white paper relates how our AI technology and semantic data and directory services can improve the performance of chatbots, agents, and assistants that interact with each other as well as with the people using them.

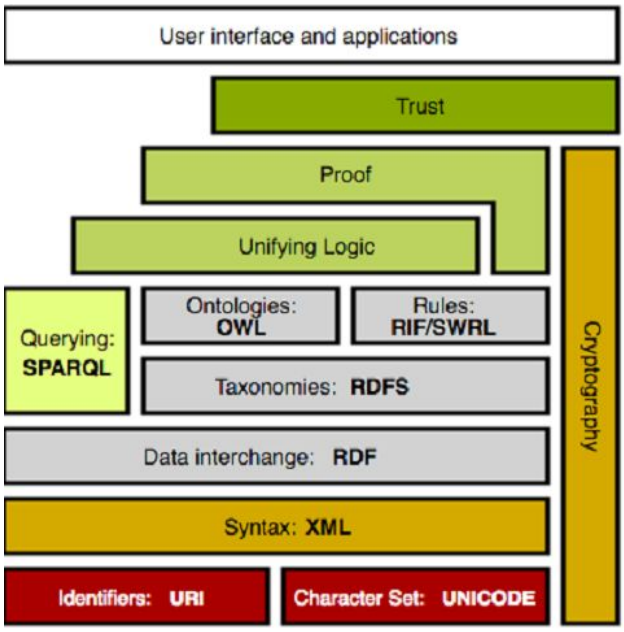
Our database of machine readable semantic data covers over 5 million concepts organized into thousands of taxonomies specifically related to the meaning of things within domains of knowledge ranging from Anime to Zoology. The database contains over 100 million topical sentences in common formats used in Intelligent Agent applications. There are 4.9 million images of unique concepts also available.

Figure 1. Concepts are the meaning of topics in a taxonomy



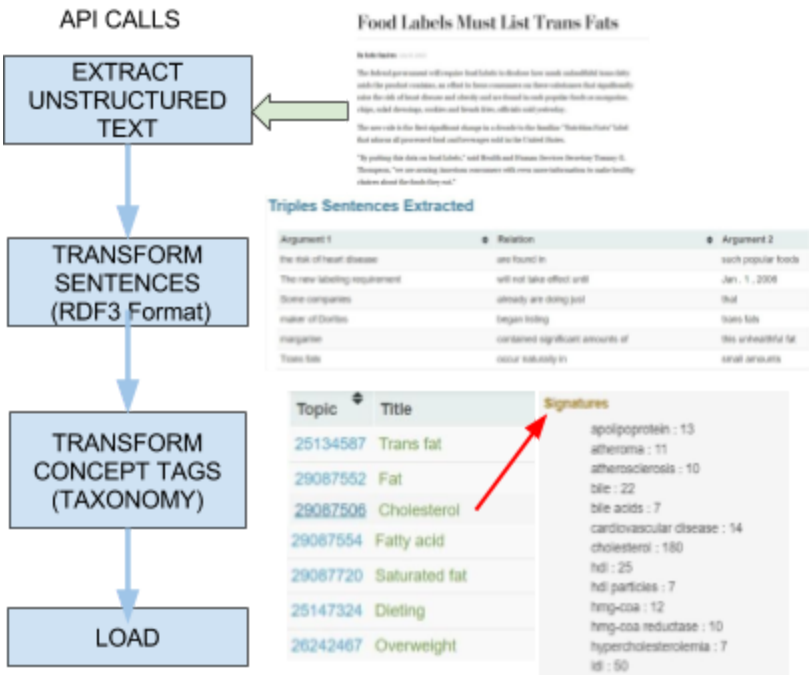
The heart of Intellisophic's technology is the identification of topics as concepts inside unstructured data (text). A concept is any abstract thought or idea. It is defined by a collection of words and phrases that distinguish one concept from another within a specific taxonomy. The taxonomy is the "context."

Figure 2. The Semantic Web Technology Stack



The Intellisophic technology stack is contained in the broader Semantic Web stack related to the Taxonomies and Data Interchange RDF sections in grey. Intellisophic stores Semantic Data on AWS in S3. Applications can access the data using the standard formats for machine interchange such as OWL, RIF and WRL. The blue links connect to Linkapedia and provide information about each topic, the largest information directory ever created. Linkapedia was built using Intellisophic APIs and technology stack RDFS and RDF.

Figure 3. The Semantic Data Extraction Process Flow



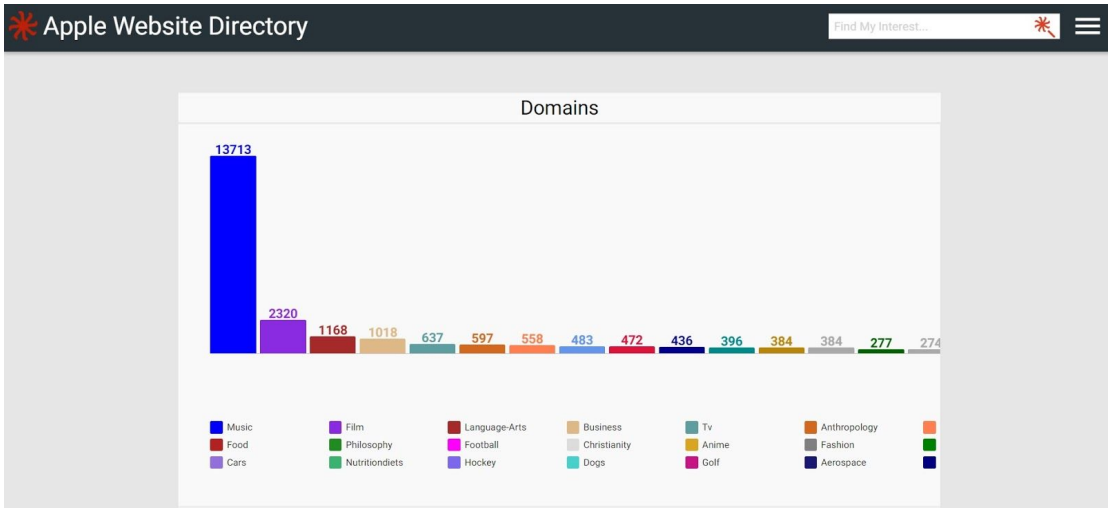
The Intellisophic process is accessed through APIs. Figure 3. shows a semantic data extraction process that starts with a text document and creates RDF triples and concept tags which can be used in Intelligent Agent Applications to model broad general knowledge.

Example: Intelligent Ad Agents

Home assistant smart speakers like Amazon Echo, Google Home, and Apple Siri and HomePod are going to become very prevalent over the next four years. There are about 20 million in homes now (Amazon has 10 million) and that could grow to hundreds of millions or more. Just as adtech has changed to reach new audiences as media evolves, this example envisions intelligent ads (IAd) that are in-native and have deep knowledge of the brand product lines expressed in taxonomies, concepts, and descriptive sentences as semantic data. The IAd would be delivered through the home assistant as an interactive intelligent sales pitch that can identify interests of the user through conversation.

Step 1: Intellisophic reads and extracts the brand product knowledge from its web site.

Over 25,000 product concepts in 70 broad subject areas with 65% of the topics in Music, Film, and Books (Language arts).



These brand related topics can be explored here: https://apple.linkapedia.com with API support to perform editing and other services in the Intellisophic tool kit.

The IAd uses this brand related semantic data and the taxonomies to engage in a sales conversation. Perhaps to offer a sale of a music album after discovering the interests of a home user in German Hip-hop. The use of taxonomies to browse interests is not search for words spoken to suggest web links.

Step 2: Use taxonomy and concepts specific to the brand to determine product interest.

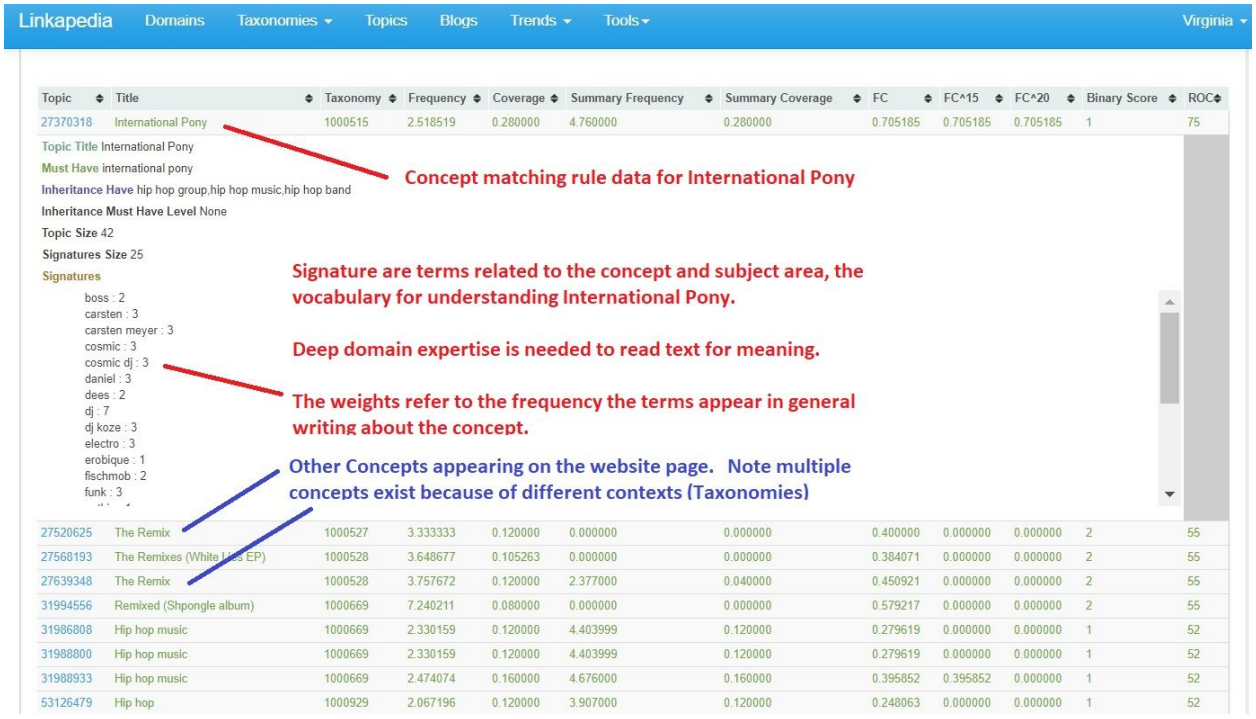
Example concept: ***International Pony*** (a german hip hop combo)
Intellisophic Taxonomy Database Entry
https://www.linkapedia-music.com/topics/hip-hop-groups-by-nationality/international-pony/81307372/preview?blog-id=e10bc12cddbffcea3ad44292171d9769

Music > Hip hop groups by nationality > German hip hop groups > International Pony

Apple website page with International Pony reference
https://itunes.apple.com/bw/album/you-saw-it-all-dj-koze-remix/id921490593?i=921490615

Concepts extracted from Apple web page:

Figure 4. Concept extraction example



Topic	Title	Taxonomy	Frequency	Coverage	Summary Frequency	Summary Coverage	FC	FC^15	FC^20	Binary Score	ROC
27370318	International Pony	1000515	2.518519	0.280000	4.760000	0.280000	0.705185	0.705185	0.705185	1	75
27520625	The Remix	1000527	3.333333	0.120000	0.000000	0.000000	0.400000	0.000000	0.000000	2	55
27568193	The Remixes (White Label EP)	1000528	3.648677	0.105263	0.000000	0.000000	0.384071	0.000000	0.000000	2	55
27639348	The Remix	1000528	3.757672	0.120000	2.377000	0.040000	0.450921	0.000000	0.000000	2	55
31994556	Remixed (Shpongle album)	1000669	7.240211	0.080000	0.000000	0.000000	0.579217	0.000000	0.000000	2	55
31986808	Hip hop music	1000669	2.330159	0.120000	4.403999	0.120000	0.279619	0.000000	0.000000	1	52
31988800	Hip hop music	1000669	2.330159	0.120000	4.403999	0.120000	0.279619	0.000000	0.000000	1	52
31988933	Hip hop music	1000669	2.474074	0.160000	4.676000	0.160000	0.395852	0.395852	0.000000	1	52
53126479	Hip hop	1000929	2.067196	0.120000	3.907000	0.120000	0.248063	0.000000	0.000000	1	52

The rich knowledge contained in the Taxonomies and concepts make it possible for the Intelligent Assistant to carry on conversations that are nearly unbounded. The usual chat bot script has both a limited vocabulary and limited branch points and often recycles.

Step 3: Use the RDF triples extracted to increase engagement by understanding and inferring information.

Natural Language Processing extracts sentences from the Apple web page and then extracted RDF triples. This example is raw output before more complex transformations were applied.

Figure 5. Data from intermediate RDF3 extractor.

Note that a single sentence can have multiple triples. The example has been unified at the predicate level. The raw predicate “was known first for his exceptional skills” was sent to a Unifier bot that mapped the text into the formal predicate hasSkill. Other bots would provide the unification of words like His, He, it to possible identifiable subjects or objects. In this example there is a logic inference related to time that was used to infer that International Pony is after year 2000.

Sentence	RDF triples
S1 “Germany native stefan kozalla was known first for his exceptional dj skills.	1. stefan kozalla: hasSkill: dj 2. stefan kozalla: isNative: German
S2 “During the '90s, he was a member of hip-hop group Fischmob and debuted later in the decade as Adolf Noise.”	1. He=stefan kozalla: isMember :Fischmob 2. Fischmob inPeriod:1990 - 1999 3. Adolf Noise inPeriod:1995 - 1999 4. He=stefan kozalla: isMember: Adolf Noise
S3 “Music Is Okay, issued in 2000 on Yo Mama, was his first mix album, released prior to the formation of International Pony -- a whimsical dance act that was signed to Columbia in Europe.”	1. Music is Okay: issueDate 2000 2. Music is Okay: issueLabel Yo Mama 3. International Pony: isAfter: 2000 4. International Pony: isSignedBy:Columbia Europe 5. International Pony: isA: whimsical dance act 6. He=stefan kozalla: createWork : mix album.

This example is about the future but the Intellisophic taxonomies and technology are available today.

About Intellisophic, A Linkapedia Company

http://www.intellisophic.com.

Intellisophic is the world's largest provider of taxonomic content. Unlike other methods for taxonomy development that are limited by the expense of corporate librarians and subject matter experts, Intellisophic content is machine developed at a fraction of the cost of manual methods. Our taxonomy library covers **over five million topics** defined by hundreds of millions of vocabulary terms.

Intellisophic provides **Concept Tagging** and **Data Enrichment** solutions for organizations with large quantities of unstructured data. Intellisophic enables solutions for content management, taxonomy enrichment & editing, directory building, business intelligence, data warehousing, and search infrastructure. Applications fueled by Intellisophic's technology and taxonomies are currently used by our global customers with applications in Pharmaceutical, content tagging, intelligence, and web portal/Reference applications.

About Linkapedia

www.linkapedia.com

Linkapedia is the largest directory of web content ever created and was built to demonstrate our semantic AI technology and commitment to improving information access across all subject areas for mobile users everywhere.

Contact:

Volker Kuebler, COO 650-810-5783 volker@intellisophic.com

volker@linkapdia.com

APPLICATION SUCCESS STORIES

Search & Retrieval

- Enterprise Search
- Next-generation Consumer Search
- Vertical Search

Categorization & Classification

- RSS - Customized news & content feeds
- Unstructured Data visualization

Business Intelligence / Data Warehousing

- Unstructured Data Profiling

Portal Infrastructure

- Categorization/search and indexing
- Faceted Search

Compliance Management / Information Security

- Outbound file monitoring (docs, email, IM, databases, etc.)

Social Networking

- Conversation, Resume profiling

Electronic Discovery

- Document forensics
- Compliance, litigation support

Government Intelligence

- Search, Categorization, Tagging, Profiling, Discovery, Data Triage, Ontology Building