

# NLP: Unlock the Hidden Business Value in Voice Communications



By Dr. Jans Aasman, CEO, Franz Inc.

Today organizations capture an enormous amount of information in spoken conversations, from routine customer service calls to sophisticated claims processing interactions in finance and healthcare. But most of this information remains hidden and unused due to the difficulty of turning these conversations into meaningful data that can be effectively analyzed through Natural Language Processing (NLP).

Simply applying speech recognition software to voice conversations often results in unreliable data. State-of-the-art speech recognition systems still have trouble distinguishing between homophones (words with the same pronunciation, but different meanings), as well as the difference between proper names (i.e. people, products) and separate words. In addition, there is also the challenge of identifying domain-specific words accurately. Thus, in most cases, using speech recognition software alone doesn't produce accurate enough data for reliable NLP.

Domain-specific taxonomies are key to understanding conversations via speech recognition systems. With them, we can feed conversations to knowledge graphs that understand the conversation and make connections in the data. Knowledge graphs provide the ability to extract the correct meaning of

text from conversations and connect concepts in order to add business value.

Knowledge graphs fed with NLP provide two prime opportunities for monetization. First, organizations can better understand their customers to improve products and services more to their liking, which in turn boosts marketing, sales and customer retention rates. Secondly, this analysis gives contact center agents real-time support for optimizing customer interactions to produce faster resolutions, better conversion rates, and cross-selling and up-selling opportunities. These approaches enable companies to capitalize on speech recognition knowledge graphs, accelerate their ROI, and expand their bottom lines.

## **Taxonomy Driven Speech Recognition**

The story of taxonomy-driven speech recognition closely relates to knowledge graphs. The first wave of knowledge graphs was built from taking structured data and turning it into semantic graphs that support the linked open data movement. The next wave is all about unstructured data. People started doing Natural Language Processing on documents and textual conversations like emails and chats. Doing so accurately for a given domain requires a taxonomy to understand the words and concepts. Otherwise, downstream processes like entity extraction and event detection won't work.

Read the full article at [DZone](#).

---

# Data-Centric Architecture Forum – DCAF 2021

Data and the subsequent knowledge derived from information are the most valuable strategic asset an organization possesses. Despite the abundance of sophisticated technology developments, most organizations don't have disciplines or a plan to enable data-centric principles.

## **DCAF 2021 will help provide clarity.**

Our overarching theme for this conference is to **make it REAL**. Real in the sense that others are becoming data-centric, it is achievable, and you are not alone in your efforts.

Join us in understanding how data as an open, centralized resource outlives any application. Once globally integrated by sharing a common meaning, internal and external data can be readily integrated, unlike the traditional “application-centric” mindset predominantly used in systems development.

The compounding problem is these application systems each have their own completely idiosyncratic data models. The net result is that after a few decades, hundreds or thousands of applications implemented have given origin to a segregated family of disparate data silos. Integration debt rises and unsustainable architectural complexity abounds with every application bought, developed, or rented (SaaS).

**Becoming data-centric will improve data characteristics** of findability, accessibility, interoperability, and re-usability (FAIR principles), thereby allowing data to be exported into any needed format with virtually free integration.\



**Dr. Jans Aasman to present –  
Franz's approach to Entity Event  
Data Modeling for Enterprise  
Knowledge Fabrics**

---

# **Sharing Ontologies Globally To Speed Science And Healthcare Solutions – OntoPortal**

## **International Ontology Sharing Is Becoming A Reality**

A consortium of researchers recently formed an organization dedicated to standardizing how scientists define their ontologies, which are essential for retrieving datasets as well as understanding and reproducing research. The group called OntoPortal Alliance is creating a public repository of internationally shared domain-specific ontologies. All the repositories will be managed with a common OntoPortal appliance that has been tested with AllegroGraph Semantic Knowledge Graph software. This enables any OntoPortal adopter to get all the power, features, maintainability, and support benefits that come from using a widely adopted, state-of-the-art semantic knowledge graph database.

Read the full article at [HealthIT Outcomes](#) –

**As Dr. Jans Aasman, CEO of Franz Inc. explains,** “When building a Knowledge Graph as your enterprise’s single source of truth, it’s critical to include ontologies and taxonomies. AI applications and complex reasoning analytics require information from both databases and knowledge bases that contain domain information, taxonomies, and ontologies to solve complex questions. To make this possible, we developed a novel hybrid sharding technology called FedShard, which facilitates the combination of data and knowledge required by applications like Montefiore’s PALM. But this approach is not unique or specific to Healthcare, it is applicable in many other industries, which is why we are excited about OntoPortal’s plans to bring sharing of domain ontologies to a broad audience.”



---

# Document Knowledge Graphs with NLP and ML

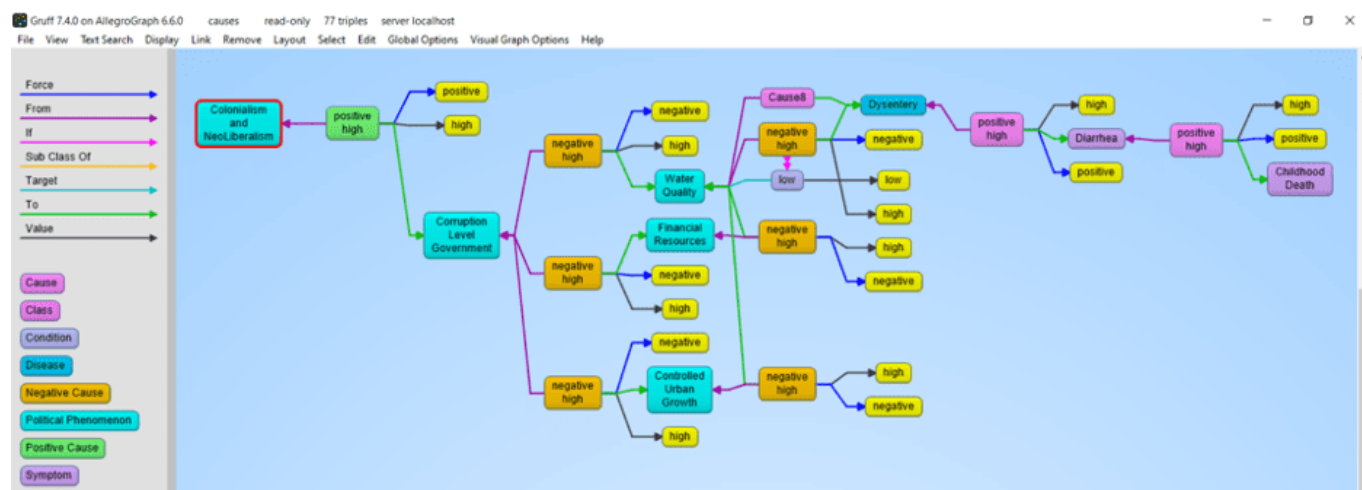
A core competency for Franz Inc is turning text and documents into Knowledge Graphs (KG) using Natural Language Processing (NLP) and Machine Learning (ML) techniques in combination with AllegroGraph. In this document we discuss how the techniques described in [NLP and ML components of AllegroGraph] can be combined with popular software tools to create a robust Document Knowledge Graph pipeline.

We have applied these techniques for several Knowledge Graphs but in this document we will primarily focus on three completely different examples that we summarize below. First is the Chomsky Legacy Project where we have a large set of very dense documents and very different knowledge sources, Second is a knowledge graph for an intelligent call center where we have to deal with high volume dynamic data and real-time decision support and finally, a large government organization where it is very important that people can do a semantic search against documents and policies that steadily change over time and where it is important that you can see the history of documents and policies.

## **Example [1] Chomsky Knowledge Graph**

The Chomsky Legacy Project is a project run by a group of admirers of Noam Chomsky with the primary goal to preserve all his written work, including all his books, papers and interviews but also everything written about him. Ultimately students, researchers, journalists, lobbyists, people from the AI community, and linguists can all use this knowledge graph for their particular goals and questions.

The biggest challenges for this project are finding causal relationships in his work using event and relationship extraction. A simple example we extracted from an author quoting Chomsky is that neoliberalism ultimately causes childhood death.



## Example 2: N3 Results and the Intelligent Call Center

This is a completely different use case (See a recent KMWorld Article <https://allegrograph.com/knowledge-graphs-enhance-customer-experience-through-speed-and-accuracy/>). Whereas the previous use case was very static, this one is highly dynamic. We analyze in real-time the text chats and spoken conversations between call center agents and customers. Our knowledge graph software provides real-time decision support to make the call center agents more efficient. N3 Results helps big tech companies to sell their high tech solutions, mostly cloud-based products and services but also helps their clients sell many other technologies and services.

The main challenge we tackle is to really deeply understand what the customer and agent are talking about. None of this can be solved by only simple entity extraction but requires elaborate rule-based and machine learning techniques. Just to give a few examples. We want to know if the agent talked about their most important talking points: that is, did the agent ask if the customer has a budget, or the authority to make a decision or a timeline about when they need the new technology

or whether they actually have expressed their need. But also whether the agent reached the right person, and whether the agent talked about the follow-up. In addition, if the customer talks about competing technology we need to recognize that and provide the agent in real-time with a battle card specific to the competing technology. And in order to be able to do the latter, we also analyzed the complicated marketing materials of the clients of N3.

### **Example 3: Complex Government Documents**

Imagine a regulatory body with tens of thousands of documents. Where nearly every paragraph has reference to other paragraphs in the same document or other documents and the documents change over time. The goal here is to provide the end-users in the government with the right document given their current task at hand. The second goal is to keep track of all the changes in the documents (and the relationship between documents) over time.

### **The Document to Knowledge Graph Pipeline**



Process Name	Input	Output
1. Custom Taxonomy Creation	Corpus Analytics, Taxonomy tool	A SKOS taxonomy containing concepts, concept hierarchy, prefLabels, altLabels.
2. Document Preparation	Documents (pdf, word, ppt, xlsx), Apache Tika, Spacy for XML cleanup	An XML version of each document
3. Extract Document Meta Data	Document + Apache Tika	JSON dictionary of the Document MetaData
4. XML-to-Triples	XML+JSON dictionary, XMLToTriples.py	Graph-based document tree with chapters, sections, and paragraphs as triples. Also includes meta data as triples
5. Entity-Extraction	Paragraphs + taxonomies + AllegroGraph Entity extract or external extractors	Concepts, persons, places, currencies. Connected to paragraphs
6. LOD Enrichment	Paragraphs + IBM Natural Language Understanding.	Concept categories and links to DBpedia and GeoNames, etc.
7. Complex Relationship and Event extraction.	Paragraphs + Taxonomy + Rules in Spacy or AllegroGraph	Complex events and relationships, References to other document sections.
8. NLP and ML	Chapters and paragraphs + all the tools described [here], but also using Spacy, Gensim, BERT, SciKit Learn.	Similarities, sentiment, query answering, smart search, text classification, word embeddings, abstracts
9. Versioning and Document tracking	Old + New document, compare.py	Old document in historic repository, new document in current, changed graph.
10. Statistical Relationships	Concepts + OddRatio.py or OddsRatio.cl	Statistical relationships between concepts.

Let us first give a quick summary in words of how we turn documents into a Knowledge Graph.

## **[1] Taxonomy Creation**

Taxonomy of all the concepts important to the business using open source or commercial taxonomy builders. An available industry taxonomy is a good starting point for additional customizations.

## **[2] Document Preparation**

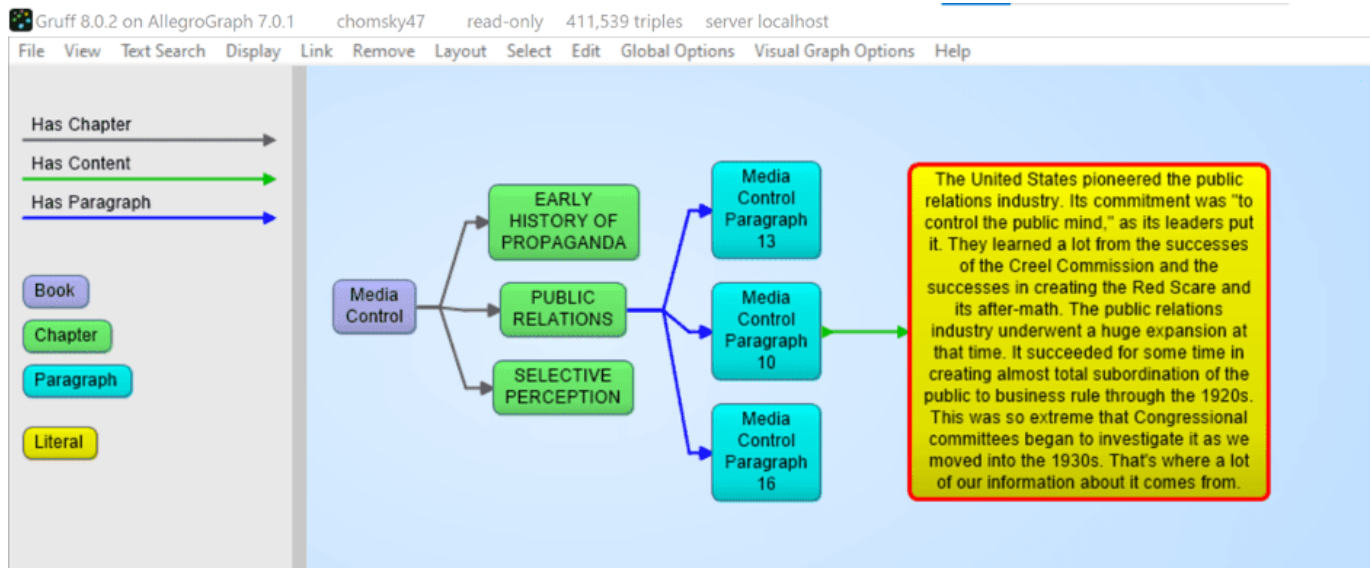
We then take a document and turn it into an intermediate XML using Apache Tika. Apache Tika supports more than 1000 document types and although Apache Tika is a fantastic tool, the output is still usually not clean enough to create a graph from, so we use Spacy rules to clean up the XML to make it as uniform as possible.

## **[3] Extract Document MetaData**

Most documents also contain document metadata (author, date, version, title, etc) and Apache Tika will also deliver the metadata for a document as a JSON object.

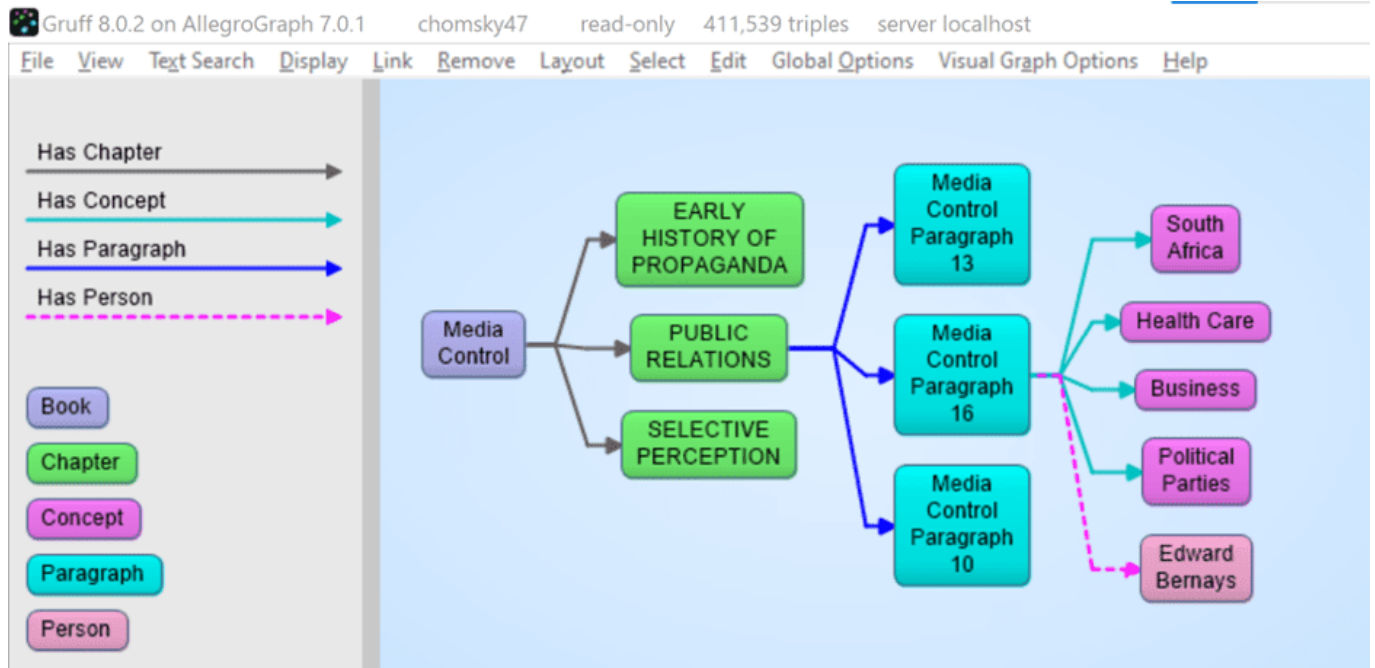
## **[4] XML to Triples**

Our tools ingest the XML and metadata and transform that into a graph-based document tree. The document is the root and from that, it branches out into chapters, optionally sections, all the way down to paragraphs. The ultimate text content is in the paragraphs. In the following example we took the XML version of Noam Chomsky's book Media Control and turned that into a tree. The following shows a tiny part of that tree. We start with the Media Control node, then we show three (of the 11) chapters, for one chapter we show three (of the 6) paragraphs, and then we show the actual text in that paragraph. We sometimes can go even deeper to the level of sentences and tokens but for most projects that is overkill.



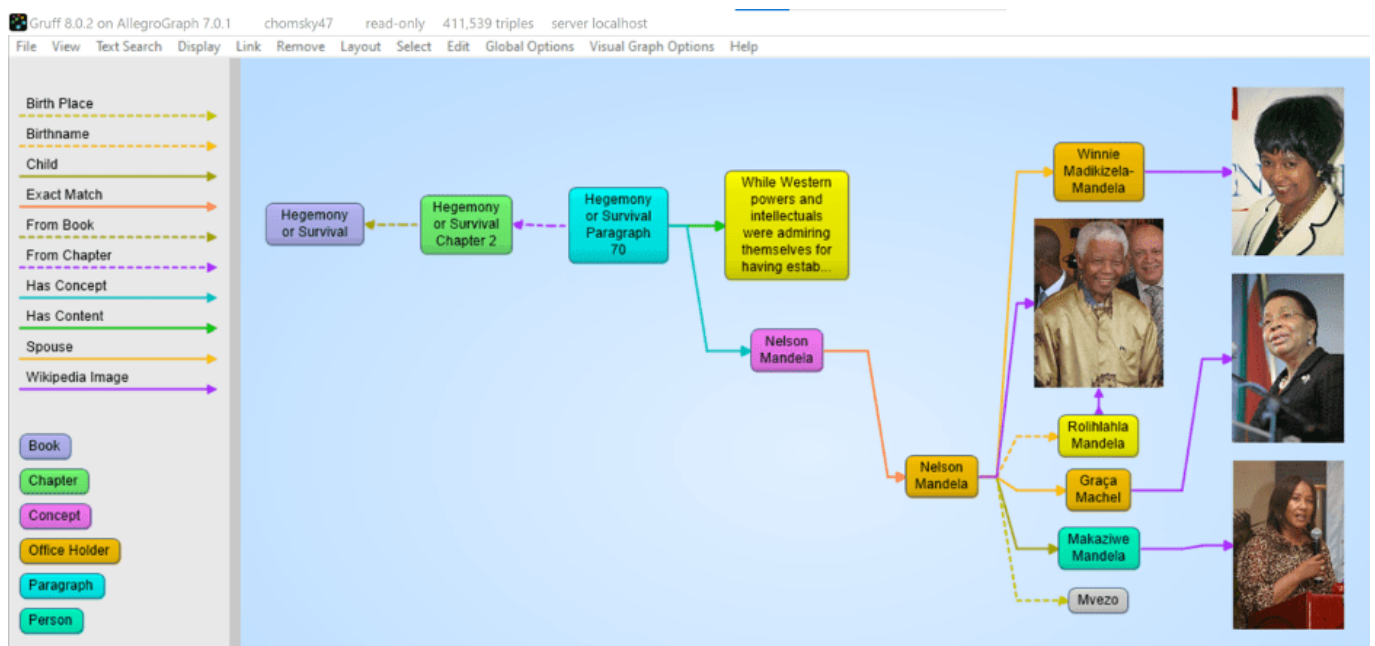
## [5] Entity Extractor

AllegroGraph's entity extractor takes as input the text of each paragraph in the document tree and one or more of the taxonomies and returns recognized SKOS concepts based on `prefLabels` and `altLabels`. AllegroGraph's entity extractor is state of the art and especially powerful when it comes to complex terms like product names. We find that in our call center a technical product name can sometimes have up to six synonyms or very specific jargon. For example the Cisco product Catalyst 9000 will also be abbreviated as the cat 9k. Instead of developing `altLabels` for every possible permutation that human beings *will* use, we have specialized heuristics to optimize the yield from the entity extractor. The following picture shows 4 (of the 14) concepts discovered in paragraph 16. Plus one person that was extracted by IBM's NLU.



## [6] Linked Data Enrichment

In many use cases, AllegroGraph can link extracted entities to concepts in the linked data cloud. The most prominent being DBpedia, wikidata, the census database, GeoNames, but also many Linked Open Data repositories. One tool that is very useful for this is IBM's Natural Language Understanding program but there are others available. In the following image we see that the Nelson Mandela entity (Red) is linked to the dbpedia entity for Nelson Mandela and that then links to the DBpedia itself. We extracted some of his spouses and a child with their pictures.



## [7] Complex Relationship and Event Extraction

Entity extraction is a first good step to 'see' what is in your documents but it is just the first step. For example: how do you find in a text whether company C1 merged with company C2. There are many different ways to express the fact that a company fired a CEO. For example: Uber got rid of Kalanick, Uber and Kalanick parted ways, the board of Uber kicked out the CEO, etc. We need to write explicit symbolic rules for this or we need a lot of training data to feed a machine learning algorithm.

## [8] NLP and Machine Learning

There are many many AI algorithms that can be applied in Document Knowledge Graphs. We provide best practices for topics like:

- [a] Sentiment Analysis, using good/bad word lists or training data.

- [b] Paragraph or Chapter similarity using statistical techniques like Gensim similarity or symbolic techniques where we just the overlap of recognized entities as a function of the size of a text.

- [c] Query answering using word2vec or more advanced techniques like BERT

- [d] Semantic search using the hierarchy in SKOS taxonomies.

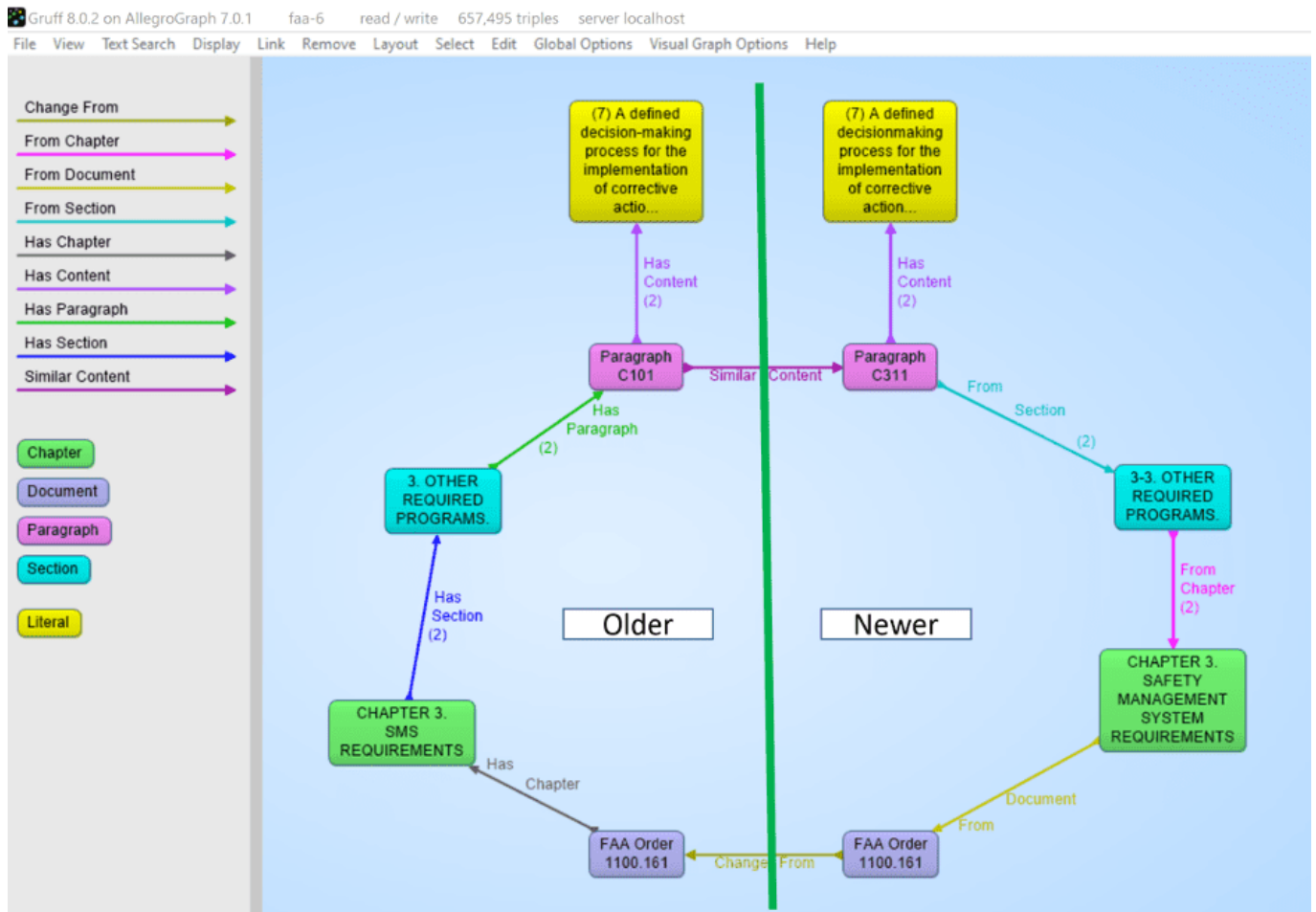
- [e] Summarization techniques for Abstractive or Extractive abstracts using Gensim or Spacy.

## **[9] Versioning and Document tracking**

Several of our customers with Document Knowledge Graphs have noted the one constant in all of these KGs is that documents change over time. As part of our solution, we have created best practices where we deal with these changes. A crucial first step is to put each document in its own graph (i.e. the fourth element of every triple in the document tree is the document id itself). When we get a new version of a document the document ID changes but the new document will point back to the old version. We then compute which paragraphs stayed the same within a certain margin (there are always changes in whitespace) and we materialize what paragraphs disappeared in the new version and what new paragraphs appeared compared to the previous version. Part of the best practice is to put the old version of a document in a historical database that at all times can be federated with the 'current' set of documents.

Note that in the following picture we see the progression of a document. On the right hand side we have a newer version of a document 1100.161 with a chapter -> section -> paragraph -> contents where the content is almost the same as the one in

the older version. But note that the newer one spells 'decision making' as one word whereas the older version said 'decision-making'. Note that also the chapter titles and the section titles are almost the same but not entirely. Also, note that the new version has a back-pointer (changed-from) to the older version.

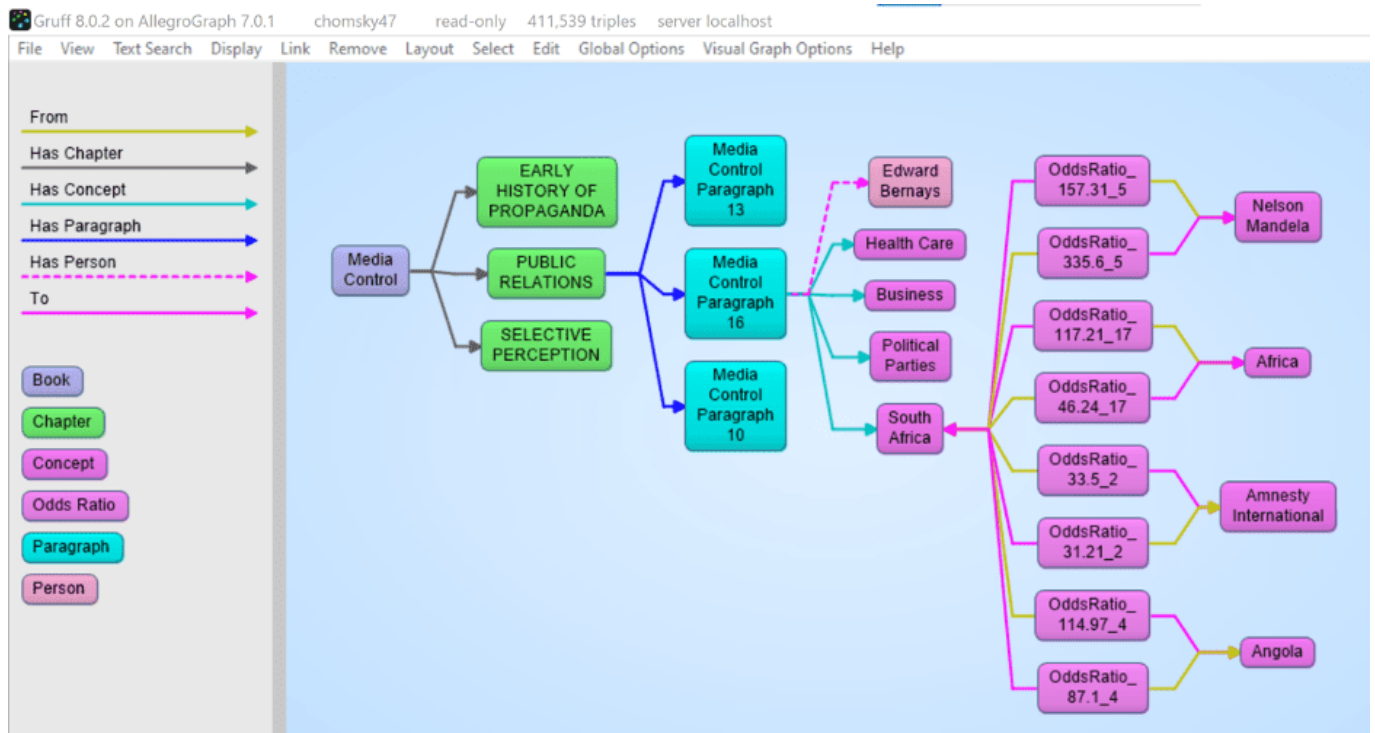


## [10] Statistical Relationships

One important analytic one can do on documents is to look at the co-occurrence of terms. Although, given that certain words might occur more frequently in text, we have to correct the co-occurrence between words for the frequency of the two terms in a co-occurrence to get a better idea of the 'surprisingness' of a co-occurrence. The platform offers several techniques in Python and Lisp to compute these co-occurrences. Note that in the following picture we computed the odds ratios between recognized entities and so we see in



the following gruff picture that if Noam Chomsky talks about South Africa then the chances are very high he will also talk about Nelson Mandela.



## Answering the Question Why: Explainable AI



The statistical branch of Artificial Intelligence has enamored organizations across industries, spurred an immense amount of capital dedicated to its technologies, and entranced numerous media outlets for the past couple of years. All of this attention, however, will ultimately prove unwarranted unless organizations, data scientists, and various vendors can answer one simple question: can they provide Explainable AI?



Although the ability to explain the results of Machine Learning models—and produce consistent results from them—has never been easy, a number of emergent techniques have recently appeared to open the proverbial ‘black box’ rendering these models so difficult to explain.

One of the most useful involves modeling real-world events with the adaptive schema of knowledge graphs and, via Machine Learning, gleaming whether they’re related and how frequently they take place together.

When the knowledge graph environment becomes endowed with an additional temporal dimension that organizations can traverse forwards and backwards with dynamic visualizations, they can understand what actually triggered these events, how one affected others, and the critical aspect of causation necessary for Explainable AI.

Read the full article at [AIthority](#).

---

## 2020 Trend Setting Products – AllegroGraph

Franz Inc. is proud to announce that it has been named to the 2020 Trend Setting Products in Data Management by Database Trends and Application Magazine.

*Database Trends and Applications (DBTA)* magazine announced its seventh annual list of trend-setting products in data management and analysis. The list, “DBTA Trend-Setting Products for 2020,” recognizes products in the marketplace that are both innovative and effective in helping customers address evolving challenges and opportunities. In all, 100

products are highlighted in the special December edition of *Database Trends and Applications* magazine and on the DBTA website, [www.dbta.com](http://www.dbta.com).

“The world of data management and analytics continues to evolve rapidly with new technologies and strategies,” remarked Thomas Hogan, Group Publisher of *Database Trends and Applications*. “Cutting through the hype and identifying products that deliver results in the real world is more important than ever. This list highlights products that are truly transformative in bringing greater agility, efficiency and innovation to market.”

“We are honored to receive this acknowledgement for our efforts in delivering Enterprise Knowledge Graph Solutions,” said Dr. Jans Aasman, CEO, Franz Inc. “In the past year, we have seen demand for Enterprise Knowledge Graphs take off across industries along with recognition from top technology analyst firms that Knowledge Graphs provide the critical foundation for artificial intelligence applications and predictive analytics. Our AllegroGraph Knowledge Graph Platform Solution offers a unique comprehensive approach for helping companies accelerate the creation of Enterprise Knowledge Graphs that deliver new value to their organization.”

---

## **Ontology Summit 2020 – Knowledge Graphs**

The Ontology Summit is an annual series of events that involves the ontology community and communities related to each year’s theme chosen for the summit. The Ontology Summit

was started by Ontolog and NIST, and the program has been co-organized by Ontolog, NIST, NCOR, NCB0, IA0A, NCO\_NITRD along with the co-sponsorship of other organizations that are supportive of the Summit goals and objectives.

Knowledge graphs, closely related to ontologies and semantic networks, have emerged in the last few years to be an important semantic technology and research area. As structured representations of semantic knowledge that are stored in a graph, KGs are lightweight versions of semantic networks that scale to massive datasets such as the entire World Wide Web. Industry has devoted a great deal of effort to the development of knowledge graphs, and they are now critical to the functions of intelligent virtual assistants such as Siri and Alexa. Some of the research communities where KGs are relevant are Ontologies, Big Data, Linked Data, Open Knowledge Network, Artificial Intelligence, Deep Learning, and many others.

Dr. Jans Aasman presented – ***“Why Knowledge Graphs Hit the Hype Cycle and What they have in common”***

Presentation Page

Presentation Slides



---

# Turn Customer Service Calls

# into Enterprise Knowledge Graphs

**Franz's CEO, Jans Aasman's recent Destination CRM article:**

The need for text analytics and speech recognition has broadened over the years, becoming more prevalent and essential in the sales, marketing, and customer service departments of various types of businesses and industries. The goal is simple for these contact center use cases: provide real-time assistance to human agents interacting with potential customers to close sales, initiate them, and increase customer satisfaction.

Until fairly recently, the rich array of unstructured data encompassing client texts, chats, and phone calls was obscured from contact centers and organizations due to the sheer arduousness of speech recognition and text analytics. When readily integrated into knowledge graphs, however, these same sources become some of the most credible for improving agent interactions and achieving business objectives.

Powered by the shrewd usage of organizational taxonomies, machine learning, natural language processing (NLP), and semantic search, knowledge graphs make speech recognition and text analytics immediately accessible, enabling real-time customer interactions that can maximize business objectives—and revenues.

## **Taxonomies**

Taxonomies are the foundation of the knowledge graph approach to rapidly conveying results of speech recognition and text analytics for timely customer interactions. Agents need three types of information to optimize customer interactions: their personas (such as an executive or a purchase department representative, for example), their reasons for contacting

them, and their industries. Taxonomies are instrumental to performing these functions because they provide a hierarchy of relevant terms to organizations.

**Read the full article at [Destination CRM](#)**

---

# Why Smart Cities Need AI Knowledge Graphs

**A linked data framework can empower smart cities to realize social, political, and financial goals.**



Smart cities are projected to become one of the most prominent manifestations of the Internet of Things (IoT). Current estimates for the emerging smart city market exceed \$40 trillion, and San Jose, Barcelona, Singapore, and many other major metropolises are adopting smart technologies.

The appeal of smart cities is binary. On the one hand, the automated connectivity of the IoT is instrumental in reducing costs associated with public expenditures for infrastructure such as street lighting and transportation. With smart lighting, municipalities only pay for street light expenses when people are present. Additionally, by leveraging options for dynamic pricing with smart parking, for example, the technology can provide new revenue opportunities.

Despite these advantages, smart cities demand extensive data management. Consistent data integration from multiple locations and departments is necessary to enable interoperability between new and legacy systems. Smart cities need granular data governance for long-term sustainability.

Finally, they necessitate open standards to future-proof their perpetual utility.

Knowledge graphs—enterprise-wide graphs which link all data assets for internal or external use—offer all these benefits and more. They deliver a uniform, linked framework for sharing data in accordance with governance protocols, are based on open standards, and exploit relationships between data for business and operational optimization. They supply everything smart cities need to realize their social, political, and financial goals. Knowledge graphs can use machine learning to reinsert the output of contextualized analytics into the technology stack, transforming the IoT's copious data into foundational knowledge to spur improved civic applications.

Read the full article at [Trajectory Magazine](#)

