

AiThority Interview with Dr. Jans Aasman



Jans Aasman, please tell us about your current role and the team / technology you handle at Franz.

As CEO of Franz Inc., I drive the overall technology vision for our Enterprise Knowledge Graph solutions and ensure our customer projects deliver the ROI results expected with graph based architectures.

Franz Inc. is composed of an expert team with skills in Graph Databases, Semantic technologies, Graph Visualization, AI, NLP and Machine Learning. Our domain knowledge encompasses large enterprises in Healthcare, Pharma, Customer Support, and Intelligence Agencies.

Our main business today revolves around AllegroGraph, a Semantic Graph platform that allows infinite data integration through a patented approach unifying all data and siloed knowledge into an Entity-Event Knowledge Graph solution that can support massive big data analytics. AllegroGraph's FedShard feature utilizes patented federated sharding capabilities that drive 360-degree insights and enable complex reasoning across a distributed Knowledge Graph. AllegroGraph is utilized by dozens of the top Fortune 500 companies worldwide.

We also offer a popular data visualization and no-code query builder called Gruff – the most advanced Knowledge Graph

visualization application on the market, which we recently integrated into Franz AllegroGraph. Gruff enables users to create visual Knowledge Graphs that display data relationships in views that are driven by the user. Ad hoc and exploratory analysis can be performed by simply clicking on different graph nodes to answer questions. Gruff's unique 'Time Machine' feature provides the capability to explore temporal context and connections within data. The visual query builder within Gruff empowers both novice and expert users to create simple to highly complex queries without writing any code.

Read the full interview at [AITHority](#).

Gartner Case Study: Entity-Event Knowledge Graph for Powering AI Solutions (Montefiore)

Gartner featured Franz's customer, Montefiore Medical Center, in a research report on Montefiore's Entity-Event Knowledge Graph:

"AI solutions are often hindered by fragmented data and siloed point solutions," according to Gartner's Chief Data and Analytics Officer Research Team. "Montefiore's data and analytics leader used semantic knowledge graphs to power its AI solutions and achieved considerable cost savings as well as improvements in timeliness and the prediction accuracy of AI models." Source: Gartner Case Study: Entity-Event Knowledge Graph for Powering AI Solutions (Montefiore) – Subscription required.

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

AllegroGraph Named to 100 Companies That Matter Most in Data

Franz Inc. Acknowledged as a Leader for Knowledge Graph Solutions

Lafayette, Calif., June 23, 2020 – Franz Inc., an early innovator in Artificial Intelligence (AI) and leading supplier of Semantic Graph Database technology for Knowledge Graph Solutions, today announced that it has been named to The 100 Companies That Matter in Data by Database Trends and Applications. The annual list reflects the urgency felt among many organizations to provide a timely flow of targeted information. Among the more prominent initiatives is the use of AI and cognitive computing, as well as related capabilities such as machine learning, natural language processing, and text analytics. This list recognizes companies based on

their presence, execution, vision and innovation in delivering products and services to the marketplace.

“We’re excited to announce our eighth annual list, as the industry continues to grow and evolve,” remarked Thomas Hogan, Group Publisher at Database Trends and Applications. “Now, more than ever, businesses are looking for ways transform how they operate and deliver value to customers with greater agility, efficiency and innovation. This list seeks to highlight those companies that have been successful in establishing themselves as unique resources for data professionals and stakeholders.”

“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 recent launch of AllegroGraph 7 with FedShard, a breakthrough that allows infinite data integration to unify all data and siloed knowledge into an Entity-Event Knowledge Graph solution will catalyze Knowledge Graph deployments across the Enterprise.”

Gartner recently released a report “How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications” and have previously stated, “The application of graph processing and graph databases will grow at 100 percent annually through 2022 to continuously accelerate data preparation and enable more complex and adaptive data science.” To that end, Gartner named graph analytics as a “Top 10 Data and Analytics Trend” to solve critical business priorities. (*Source: Gartner, Top 10 Data and Analytics Trends, November 5, 2019*).

“Graph databases and knowledge graphs are now viewed as a must-have by enterprises serious about leveraging AI and predictive analytics within their organization,” said Dr. Aasman “We are working with organizations across a broad range of industries to deploy large-scale, high-performance Entity-Event Knowledge Graphs that serve as the foundation for AI-driven applications for personalized medicine, predictive call centers, digital twins for IoT, predictive supply chain management and domain-specific Q&A applications – just to name a few.”

Forrester Shortlists AllegroGraph

AllegroGraph was shortlisted in the February 3, 2020 Forrester Now Tech: Graph Data Platforms, Q1 2020 report, which recommends that organizations “Use graph data platforms to accelerate connected-data initiatives.” Forrester states, “You can use graph data platforms to become significantly more productive, deliver accurate customer recommendations, and quickly make connections to related data.”

Bloor Research covers AllegroGraph with FedShard

Bloor Research Analyst, Daniel Howard noted “With the 7.0 release of AllegroGraph, arguably the most compelling new capability is its ability to create what Franz refers to as “Entity-Event Knowledge Graphs” (or EEKGs) via its patented FedShard technology.” Mr. Howard goes on to state “Franz clearly considers this a major release for AllegroGraph. Certainly, the introduction of an explicit entity-event graph is not something I’ve seen before. The newly introduced text to speech capabilities also seem highly promising.”

AllegroGraph Named to KMWorld’s 100 Companies That Matter in Knowledge Management

AllegroGraph was also recently named to KMWorld’s 100 Companies That Matter in Knowledge Management. The KMWorld 100 showcases organizations that are advancing their products

and capabilities to meet changing requirements in Knowledge Management.

Franz Knowledge Graph Technology and Services

Franz's Knowledge Graph Solution includes both technology and services for building industrial strength Entity-Event Knowledge Graphs based on best-of-class tools, products, knowledge, skills and experience. At the core of the solution is Franz's graph database technology, AllegroGraph with FedShard, which is utilized by dozens of the top F500 companies worldwide and enables businesses to extract sophisticated decision insights and predictive analytics from highly complex, distributed data that cannot be uncovered with conventional databases.

Franz delivers the expertise for designing ontology and taxonomy-based solutions by utilizing standards-based development processes and tools. Franz also offers data integration services from siloed data using W3C industry standard semantics, which can then be continually integrated with information that comes from other data sources. In addition, the Franz data science team provides expertise in custom algorithms to maximize data analytics and uncover hidden knowledge.

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

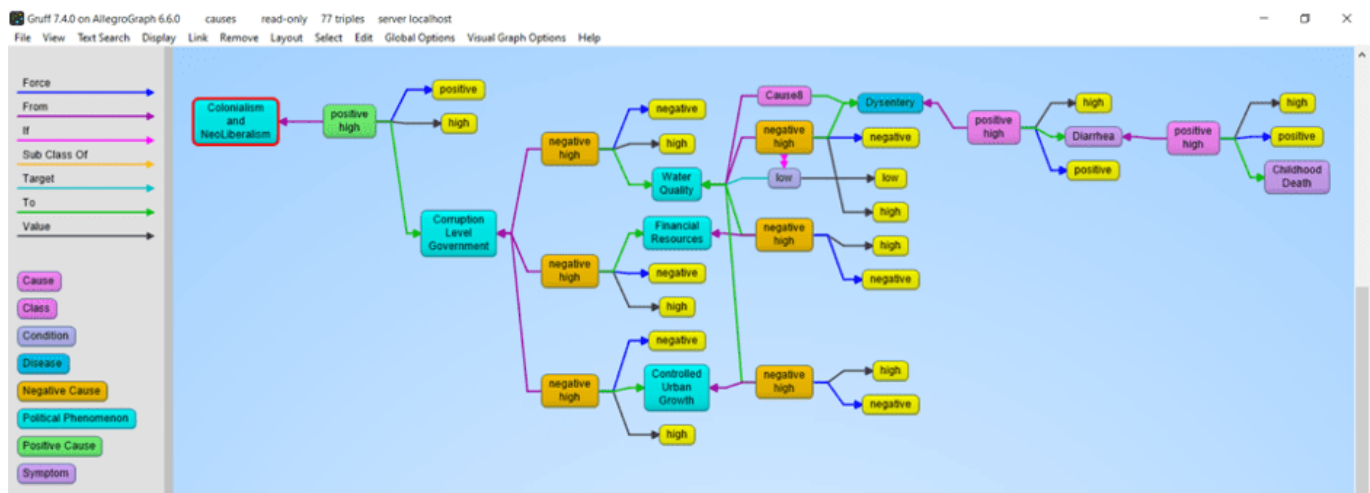
(NLP) and Machine Learning (ML) techniques in combination with AllegroGraph. In this document we discuss how the techniques described in [NLP and ML components of AllegroGraph] can be combined with popular software tools to create a robust Document Knowledge Graph pipeline.

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

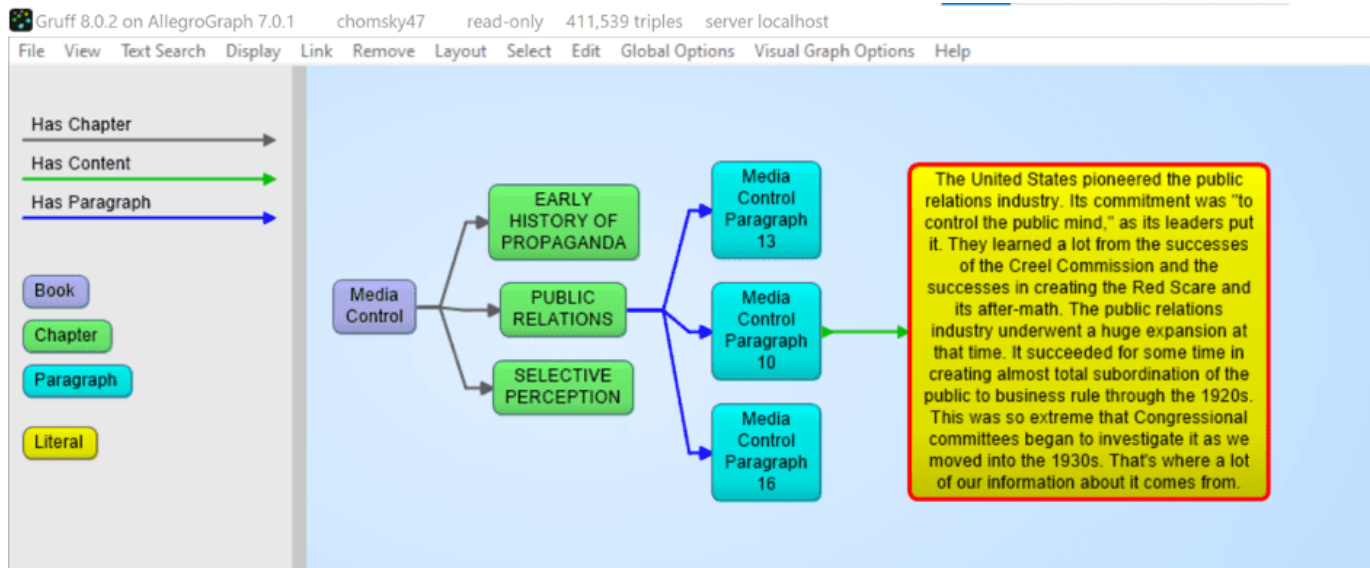
We then take a document and turn it into an intermediate XML using Apache Tika. Apache Tika supports more than 1000 document types and although Apache Tika is a fantastic tool, the output is still usually not clean enough to create a graph from, so we use Spacy rules to clean up the XML to make it as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

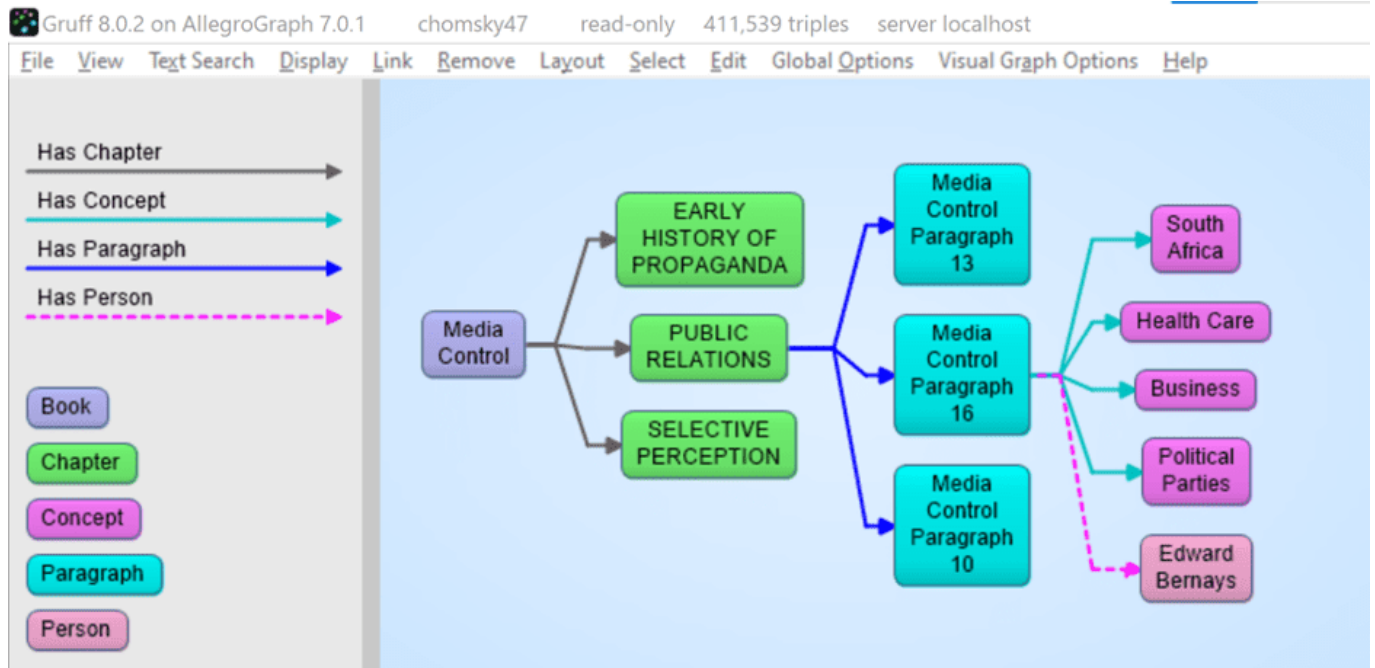
[4] XML to Triples

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

[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.

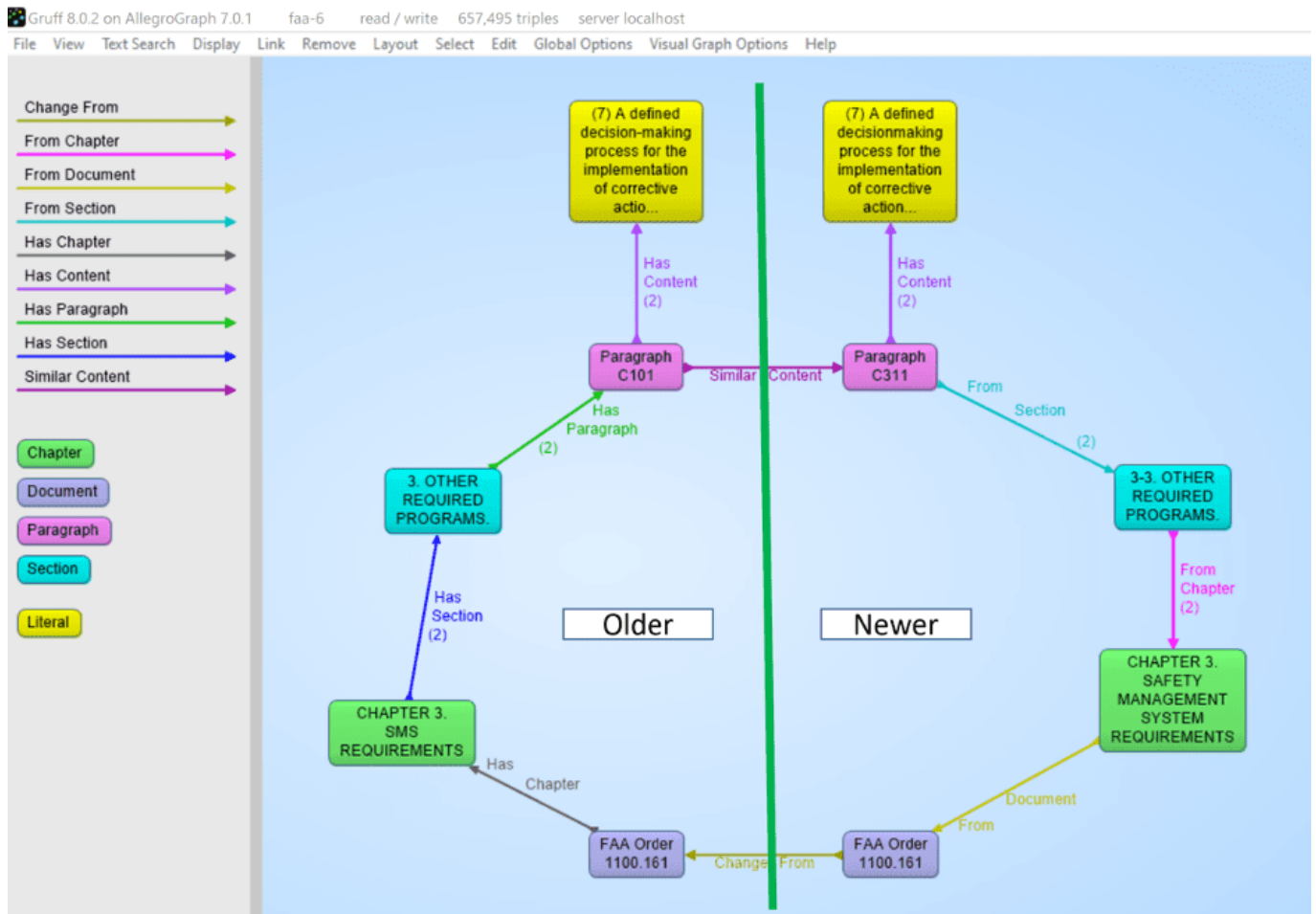
- [e] Summarization techniques for Abstractive or Extractive abstracts using Gensim or Spacy.

[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

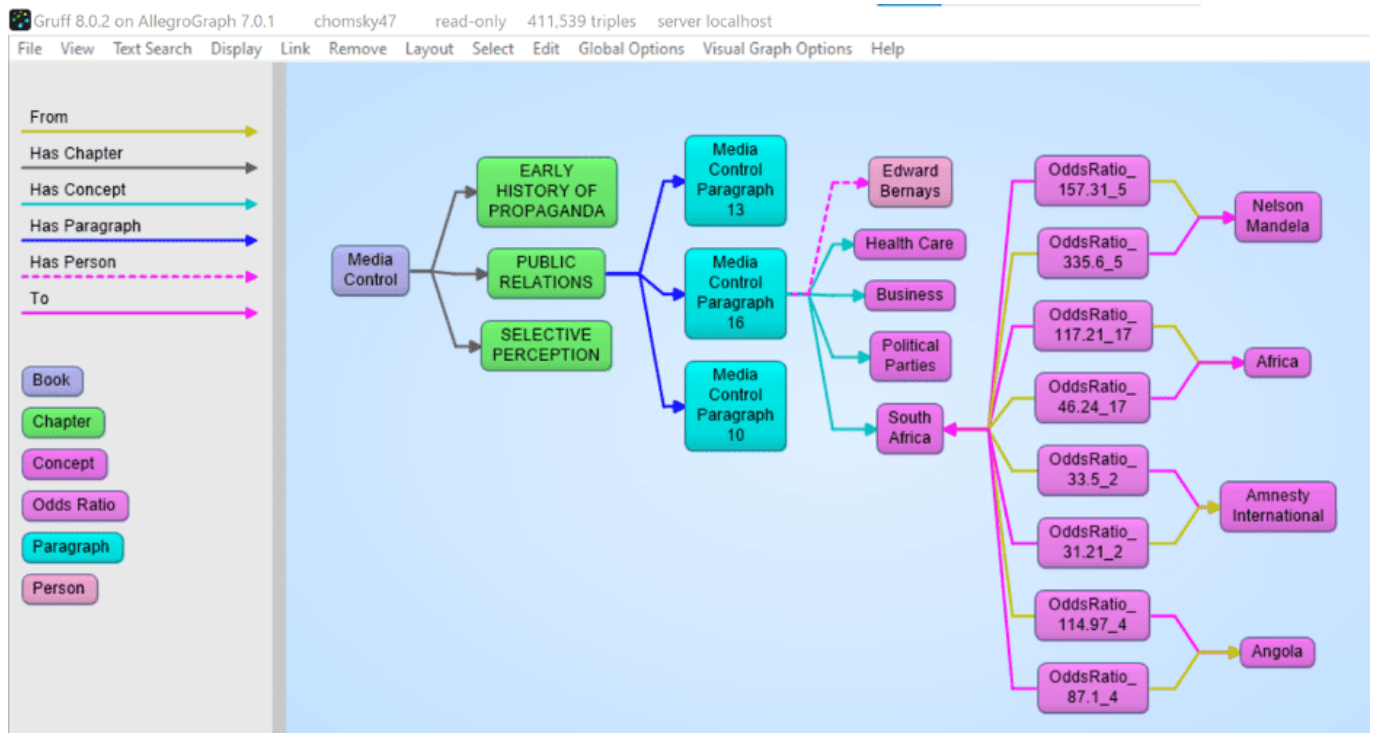
the older version. But note that the newer one spells 'decision making' as one word whereas the older version said 'decision-making'. Note that also the chapter titles and the section titles are almost the same but not entirely. Also, note that the new version has a back-pointer (changed-from) to the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



How To Avoid Another AI Winter

Forbes published the following article by Dr. Jans Aasman, Franz Inc.'s CEO.



Photo: Getty

Although there has been great progress in artificial intelligence (AI) over the past few years, many of us remember the AI winter in the 1990s, which resulted from overinflated promises by developers and unnaturally high expectations from end

users. Now, industry insiders, such as Facebook head of

AI Jerome Pesenti, are predicting that AI will soon hit another wall—this time due to the lack of semantic understanding.

“Deep learning and current AI, if you are really honest, has a lot of limitations,” said Pesenti. “We are very, very far from human intelligence, and there are some criticisms that are valid: It can propagate human biases, it’s not easy to explain, it doesn’t have common sense, it’s more on the level of pattern matching than robust semantic understanding.”



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

Franz's 2020 Predictions in the News

Looking to the future of AI, KnowledgeGraph and Semantics we had a number of publications cover our views of where AllegroGraph is headed.

Datanami

20 AI Predictions for 2020

We're still in the midst of a fake news crisis, and with the emergence of deep fakes, it will likely get worse. Luckily, we have the technology available to begin to address it, says Dr.

Jans Aasman, the CEO of Franz.

“Knowledge graphs, in combination with deep learning, will be used to identify photos and video that have been altered by superimposing existing images and videos onto source images,” Aasman says. “Machine learning knowledge graphs will also unveil the origin of digital information that has been published by a foreign source. Media outlets and social networks will use AI knowledge graphs as a tool to determine whether to publish information or remove it.”

DBTA

Ten Predictions for AI and Machine Learning in 2020

AI Knowledge Graphs will Debunk Fake News: “Knowledge Graphs in combination with deep learning will be used to identify photos and video that have been altered by superimposing existing images and videos onto source images. Machine learning knowledge graphs will also unveil the origin of digital information that has been published by a foreign source. Media outlets and social networks will use AI Knowledge Graphs as a tool to determine whether to publish information or remove it.” – Dr. Jans Aasman, CEO of Franz, Inc.

SD Times

Software predictions for 2020 from around the industry

Jans Aasman, CEO of Franz, Inc.

Digital immortality will emerge: We will see digital immortality emerge in 2020 in the form of AI digital personas for public figures. The combination of Artificial Intelligence and Semantic Knowledge Graphs will be used to transform the works of scientists, technologists, politicians and scholars into an interactive response system that uses the person’s

actual voice to answer questions. AI digital personas will dynamically link information from various sources – such as books, research papers and media interviews – and turn the disparate information into a knowledge system that people can interact with digitally. These AI digital personas could also be used while the person is still alive to broaden the accessibility of their expertise.

Dataversity

Semantic Web and Semantic Technology Trends in 2020

“The big-name Silicon Valley companies (LinkedIn, Airbnb, Apple, Uber) are all building knowledge graphs. But more importantly, Fortune 500 companies, especially banks, are also investing in knowledge graph solutions.”

IoT gets into the picture too. Aasman points to “digital twins,” which can be thought of as specialized knowledge graphs, as an exceptionally lucrative element of the technology with an applicability easily lending itself to numerous businesses. Its real-time streaming data, simulation capabilities, and relationship awareness may well prove to be the ‘killer app’ that takes the IoT mainstream, he said. As an example, by consuming data transmitted by IoT sensors, digital twins will inform the monitoring, diagnostics, and prognostics of power grid assets to optimize asset performance and utilization in near real-time.

InsideBigData

2020 Trends in Data Modeling: Unparalleled Advancement

Shapes Constraint Language (SHACL): SHACL is a framework that assists with data modeling by describing the various shapes of data in knowledge graph settings, which produces the desirable

downstream effect of enabling organizations to automate “the validation of your data,” remarked Franz CEO Jans Aasman. SHACL operates at a granular level involving classifications and specific data properties.

Workflow

2020 Trends in CyberSecurity

Software-defined perimeter transmissions also guard information at the data layer by utilizing Datagram Transport Layer Security (DTLS) encryption and Public Key Authentication. Fortifying information assets at the data layer is likely the most dependable method of protecting them, because it's the layer in which the data are actually stored. It's important to distinguish data layer security versus access layer security. The latter involves a process known as security filtering in which, based on particular roles or responsibilities, users can access data. “You can specify filters where for a particular user or a particular role whether you could see or not see particular [data],” Franz CEO Jans Aasman said. “You could say if someone has the role administrator, we're telling the system ‘administrators cannot see [certain data]’.”

Moreover, triple attributes can be based on compliance needs specific to regulations – which is immensely utilitarian in the post-GDPR data landscape. “For the government you could have a feature of whether you're a foreigner or not,” Aasman said. “HIPAA doesn't care whether you're a foreigner or not, but you can do a separate mechanism for it.”

The Importance of FAIR Data in Earth Science

Franz's CEO, Jans Aasman's recent Marine Technology News:

Data's valuation as an enterprise asset is most acutely realized over time. When properly managed, the same dataset



supports a plurality of use cases, becomes almost instantly available upon request, and is exchangeable between departments or organizations to systematically increase its yield with each deployment.

These boons of leveraging data as an enterprise asset are the foundation of GO FAIR's Findable Accessible Interoperable Reusable (FAIR) principles profoundly impacting the data management rigors of geological science. Numerous organizations in this space have embraced these tenets to swiftly share information among a diversity of disciplines to safely guide the stewardship of the earth.

According to Dr. Annie Burgess, Lab Director of Earth Science Information Partners (ESIP), the "most pressing global challenges cannot be solved by a single organization. Scientists require data collected across multiple disciplines, which are often managed by many different agencies and institutions." As numerous members of the earth science community are realizing, the most effectual means of managing those disparate data according to FAIR principles is by utilizing the semantic standards underpinning knowledge

graphs.

Read the full article at [Marine Technology News](#)

Webcast – Speech Recognition, Knowledge Graphs, and AI for Intelligent Customer Operations – April 3, 2019

Presenters – Burt Smith, N3 Results and Jans Aasman, Franz Inc.

In the typical sales organization the contents of the actual chat or voice conversation between agent and customer is a black hole. In the modern Intelligent Customer Operations center (e.g. N3 Results – www.n3results.com) the interactions between agent and customer are a source of rich information that helps agents to improve the quality of the interaction in real time, creates more sales, and provides far better analytics for management.

Join us for this Webinar where we describe a real world Intelligent Customer Operations center that uses graph based technology for taxonomy driven entity extraction, speech recognition, machine learning and predictive analytics to improve quality of conversations, increase sales and improve business visibility.

[View the recorded webinar.](#)

What is the Answer to AI Model Risk Management?

Algorithm-XLab – March 2019

Franz CEO Dr. Jans Aasman Explains how to manage AI Modelling Risks.

AI model risk management has moved to the forefront of contemporary concerns for statistical Artificial Intelligence, perhaps even displacing the notion of ethics in this regard because of the immediate, undesirable repercussions of tenuous machine learning and deep learning models.

AI model risk management requires taking steps to ensure that the models used in artificial applications produce results that are unbiased, equitable, and repeatable.



The objective is to ensure that given the same inputs, they produce the same outputs.

If organizations cannot prove how they got the results of AI risk models, or have results that are discriminatory, they are subject to regulatory scrutiny and penalties.

Strict regulations throughout the financial services industry in the United States and Europe require governing, validating, re-validating, and demonstrating the transparency of models

for financial products.

There's a growing cry for these standards in other heavily regulated industries such as healthcare, while the burgeoning Fair, Accountable, Transparent movement typifies the horizontal demand to account for machine learning models' results.

AI model risk management is particularly critical in finance.

Financial organizations must be able to demonstrate how they derived the offering of any financial product or service for specific customers.

When deploying AI risk models for these purposes, they must ensure they can explain (to customers and regulators) the results that determined those offers.

Read the full article at [Algorithm-XLab](#).