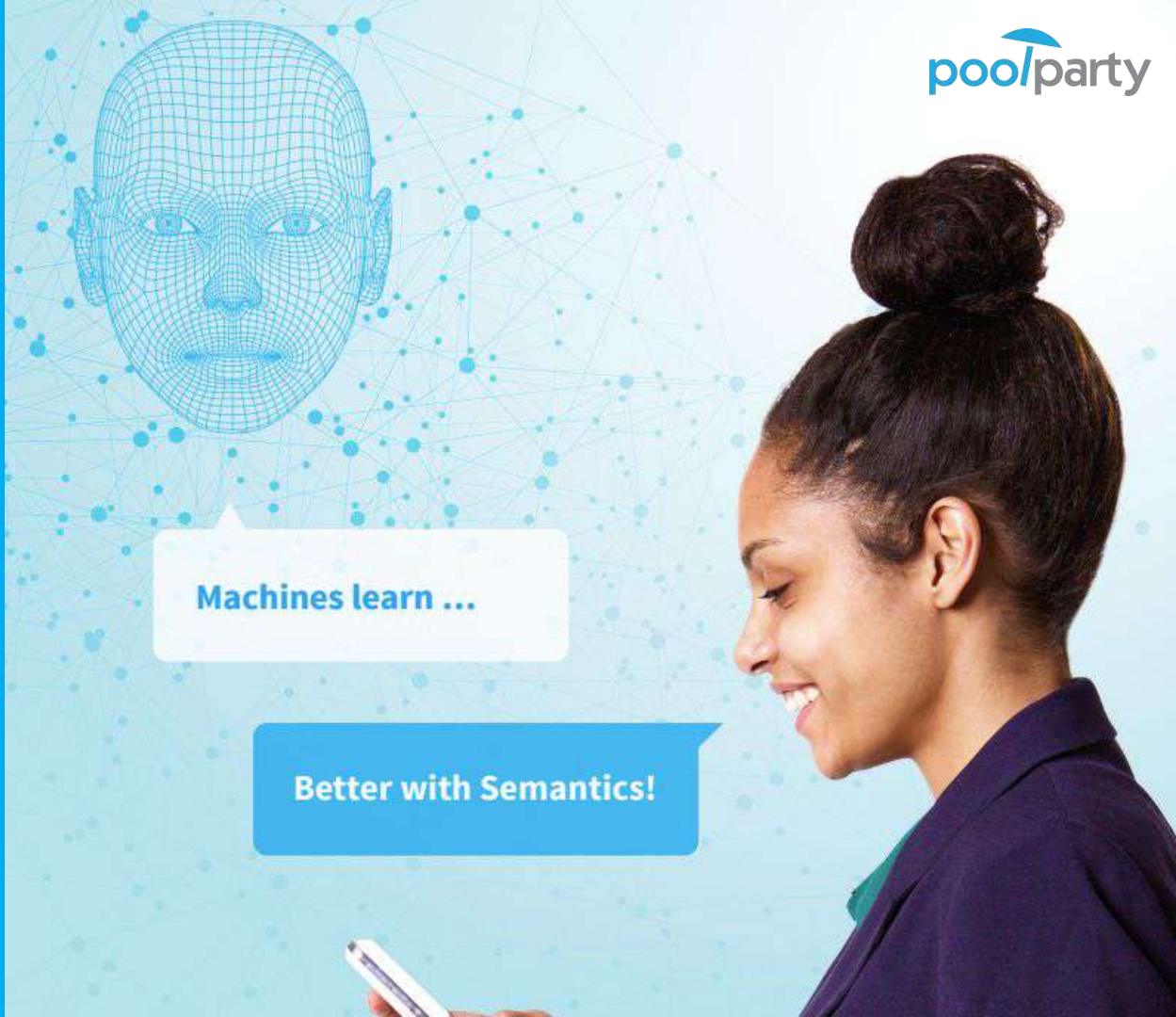


SEMANTiCS

Vienna 2018



Semantic AI for Legal Experts
Bringing Machine Learning, NLP
and Knowledge Graphs together

Andreas Blumauer
CEO & Managing Partner

**Semantic Web Company /
PoolParty Semantic Suite**

Machines learn ...

Better with Semantics!

2

Introduction

Semantic Web Company (SWC)

- ▶ Founded in 2004, based in Vienna
- ▶ Privately held

- ▶ 50 FTE
- ▶ Software Engineers & Consultants for NLP, Semantics and Machine learning

- ▶ Developer & Vendor of PoolParty Semantic Suite
- ▶ Participating in projects with €2.5 million funding for R&D
- ▶ ~30% revenue growth/year

- ▶ SWC named to KMWorld's '100 Companies That Matter in Knowledge Management' in 2016, 2017 and 2018

- ▶ www.semantic-web.com

PoolParty Semantic Suite

- ▶ Most complete Semantic Middleware on the Global Market
- ▶ *Semantic AI*: Fusion of Knowledge Graphs, NLP, and Machine Learning
- ▶ W3C standards compliant

- ▶ First release in 2009
- ▶ Current version 7.0
- ▶ On-premises or cloud-based
- ▶ Over 200 installations world-wide

- ▶ Named as Sample Vendor in Gartner's Hype Cycle for AI 2018

- ▶ KMWorld listed PoolParty as Trend-Setting Product 2015, 2016, 2017, and 2018

- ▶ www.poolparty.biz

3

SEMANTIC ARTIFICIAL INTELLIGENCE

#SemanticAI: Bringing Machine Learning,
NLP and Knowledge Graphs together

4

AI suffers from a lack of common sense, but ...



Google Assistant (2018)

5

... is talented in solving isolated problems based on isolated data sets

Face recognition



Deep Learning
Genetic Algorithms



Game AI



Monte Carlo TS
Deep Learning



Fraud detection



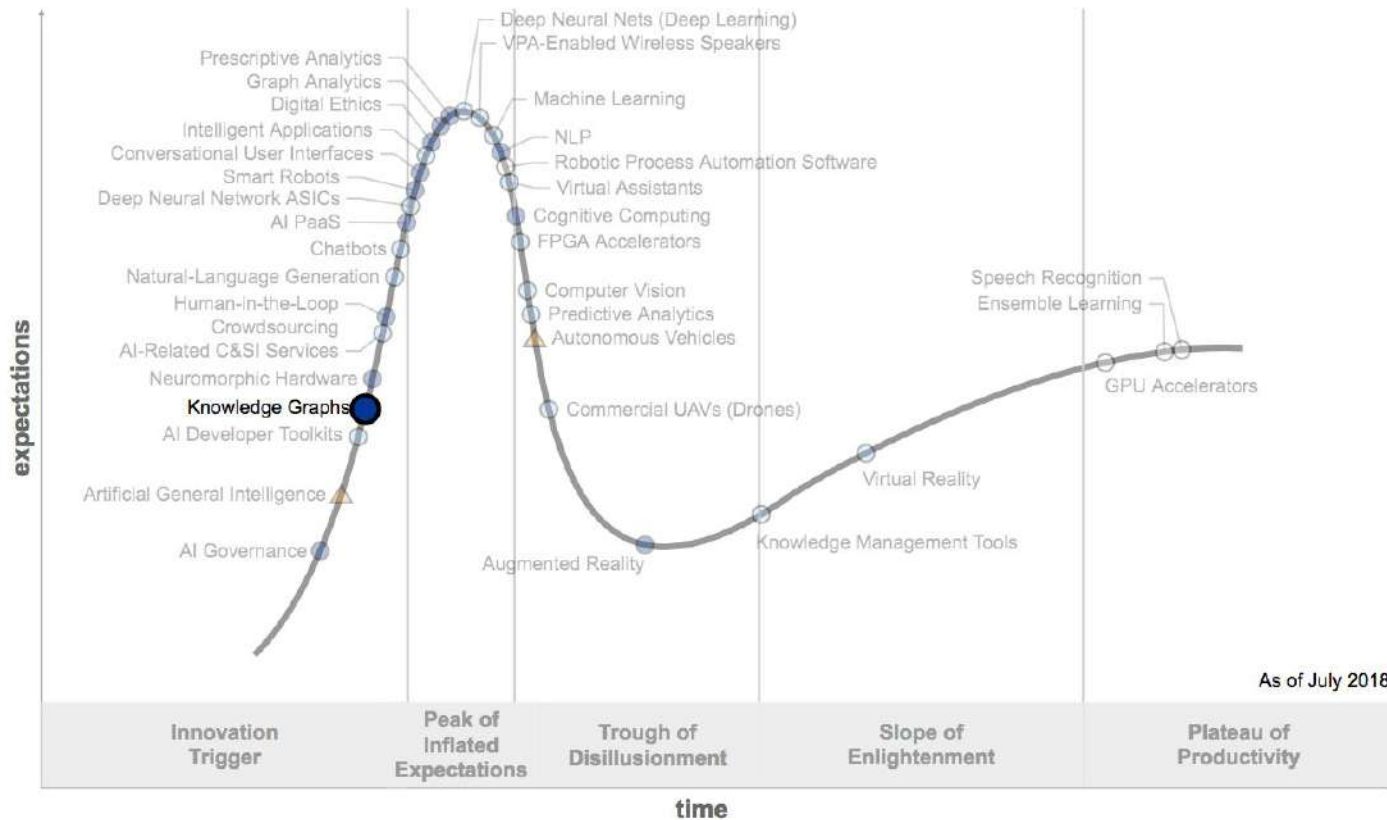
Neuronal networks
Case based reasoning



6

Gartner Hype Cycle for Artificial Intelligence, 2018

“The rising role of content and context for delivering insights with AI technologies, as well as recent knowledge graph offerings for AI applications have pulled knowledge graphs to the surface.”



As of July 2018

Plateau will be reached:
○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ✗ obsolete before plateau

7

What makes someone an intelligent being?

Assessment of the current status of Artificial Intelligence

Bloom's Taxonomy: Classify cognitive processes

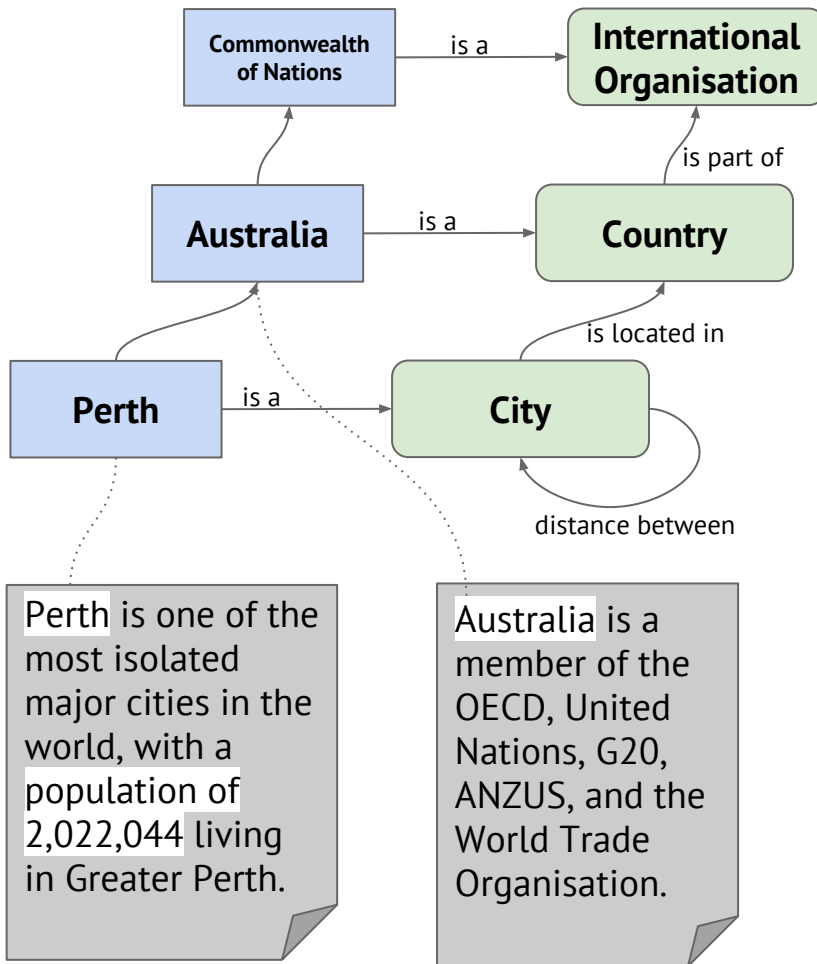
Level	Example	Questions
(6) Create	How to convert an inefficient AI system architecture to a more efficient one by replacing your choice of components?	How would you improve ...? Can you formulate a theory for ...? Can you predict the outcome if ...?
(5) Evaluate	Which kinds of knowledge models are best for machine learning, and why?	What is your opinion of ...? How would you prioritize ...? What would you use to support the view ...?
(4) Analyse	How does a graph database and a semantic knowledge model work together?	How is ... related to ...? What is the function of ...? What conclusions can you draw ...?
(3) Apply	How can taxonomies be used to enhance machine learning?	Why is ... significant? How is ... an example of ...? What elements would you use to change ...?
(2) Understand	What is the difference between an ontology and a taxonomy?	What is the difference between ...? What is the main idea of ...? Which statements support ...?
(1) Remember	Who is the inventor of the World Wide Web?	Who is ...? Where is ...? Why did ...?

8

Remember

“Knowledge graphs silently accrue ‘smart data’ — i.e., data that can be easily read and ‘understood’ by AI systems.”

Gartner Hype Cycle for Artificial Intelligence, 2018



Support complex Q&A:

Which cities located in the Commonwealth of Nations have a population of more than 2 mio. people?

Avoid illogical answers:

How far am I away from Perth, Australia?



1,195 miles

Distance between Perth and Australia

9

Six Core Aspects of Semantic AI

> [Read more](#)

1. Data Quality

Semantically enriched data serves as a basis for better data quality and provides more options for feature extraction.

2. Data as a Service

Linked data based on W3C Standards can serve as an enterprise-wide data platform and helps to provide training data for machine learning in a more cost-efficient way.

3. No black-box

Semantic AI ultimately leads to AI governance that works on three layers: technically, ethically, and on the legal layer.

4. Hybrid approach

Semantic AI is the combination of methods derived from symbolic AI and statistical AI. It is not only focused on process automation, but also on intelligence augmentation.

5. Structured data meets text

Most machine learning algorithms work well either with text or with structured data. Semantic AI is based on entity-centric data models.

6. Towards self optimizing machines

ML can help to extend knowledge graphs, and in return, knowledge graphs can help to improve ML algorithms.

10

Data Quality

Training data is semantically enriched with help from semantic knowledge models

The screenshot displays the PoolParty Semantic Classifier interface. On the left, a tree view shows 'Training Buckets' including 'Politics and conflicts', 'Crime and law', and 'Disasters and accidents', along with 'Classifiers' for 'news channel', 'news channel-cs2', 'news channel-cs', and 'news channel-c'. The main panel is titled 'news channel' and shows a 'Classifier Configuration' dropdown menu with options: Logistic Regression, Linear Support Vector Machine, Decision Tree, Gradient Boosted Tree, Deep Learning (MLP), Naive Bayes, and Random Forest. Below this, the 'Features' section includes 'Terms' (checked), 'Concepts', and 'Shadow_Concepts'. The 'Labels' section shows 'Used Labels' for 'politics', 'disasters', and 'crime'. Performance metrics include a calculated performance of 88.78% and cross-validation results: 72.55% (f1) / 72.93% (r) / 73.07% (p). Buttons for 'Train', 'Delete', 'Upload Documents', and 'Duplicate' are at the bottom.

PoolParty Semantic Classifier combines machine learning algorithms (SVM, Deep Learning, Naive Bayes, etc.) with Semantic Knowledge Graphs.

11

Benchmarking the PoolParty Semantic Classifier

Improvement of 5.2% compared to traditional (term-based) SVM



Reegle thesaurus

A comprehensive SKOS taxonomy for the clean energy sector

(<http://data.reeep.org/thesaurus/guide>)

- 3,420 concepts
- 7,280 labels (English version)
- 9,183 relations (broader/narrower + related)

Document Training Set

1,800 documents in 7 classes

Renewable Energy, District Heating Systems, Cogeneration, Energy Efficiency, Energy (general), Climate Protection, Rural Electrification

Features used	Classifier	F1 (5-fold)	Variance
Terms	LinearSVC	0.83175	0.0008
Concepts from REEGLE + Shadow Concepts	LinearSVC	0.84451	0.0011
Concepts from REEGLE	LinearSVC	0.84647	0.0009
Terms + Concepts from REEGLE + Shadow Concepts	LinearSVC	0.87474	0.0009

12

No black-box

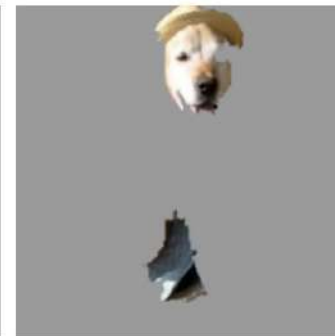
Explainable AI
(XAI)

Explainable AI as an AI whose decision-making mechanism for a specific problem **can be understood by humans who have expertise** in making decisions for that specific problem.

Explainable AI has been used for years in AI that are based on transparent methods. These include Expert Systems or Symbolic Reasoning Systems - anything that is considered GOFAI (Good Old-Fashioned AI methods)



(a) Original Image

(b) Explaining *Electric guitar*(c) Explaining *Acoustic guitar*(d) Explaining *Labrador*

Explaining an image classification prediction made by Google's Inception network, highlighting positive pixels. The top 3 classes predicted are "**Electric Guitar**" ($p = 0.32$), "**Acoustic guitar**" ($p = 0.24$) and "**Labrador**" ($p = 0.21$)

13

No black-box

Explainable AI
(XAI)

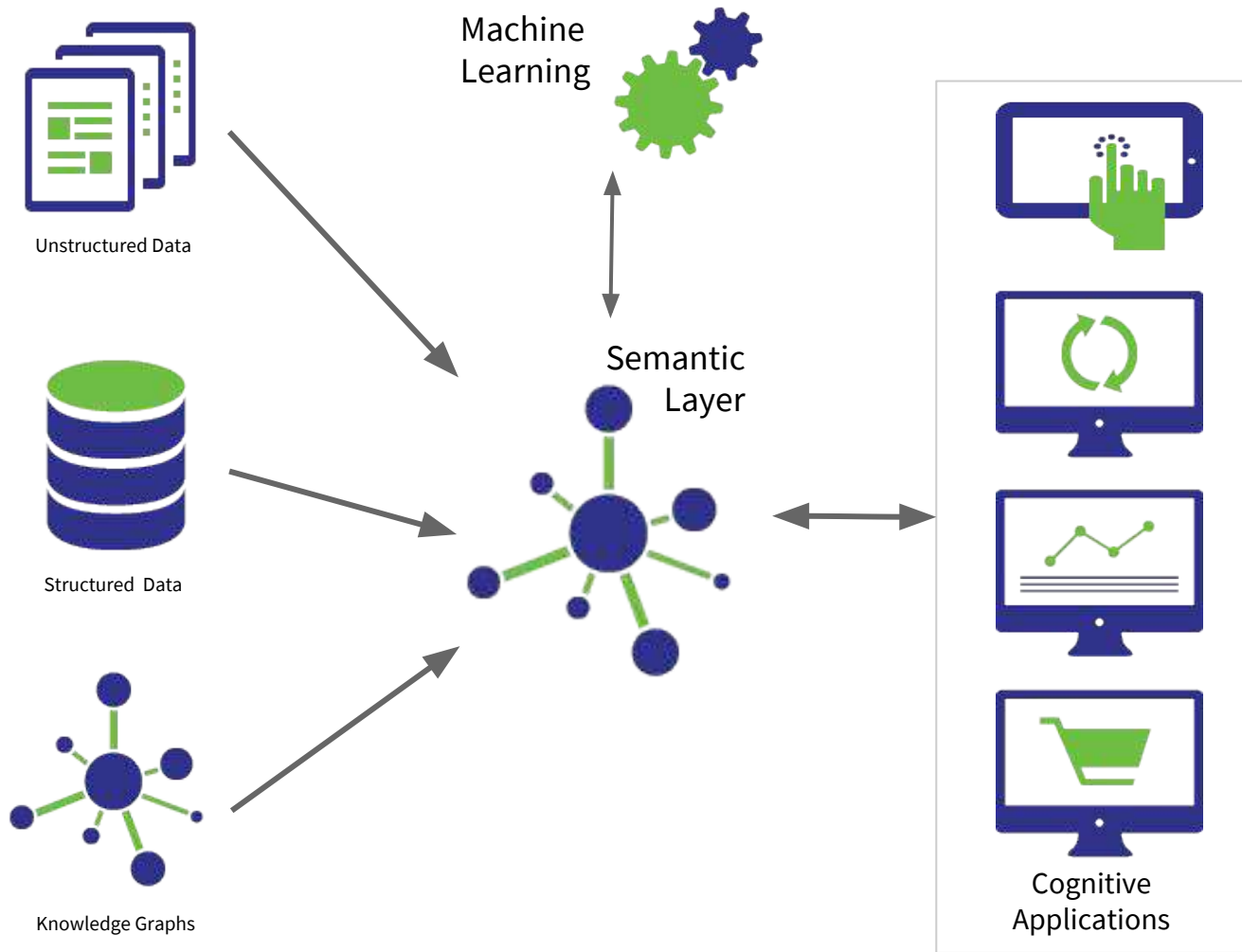
The screenshot shows the PoolParty Semantic Suite interface. On the left is a navigation tree with categories like 'Training Boxes', 'Train Classifiers', and 'Test Classifiers'. Under 'Train Classifiers', 'SVM based on Terms and Concepts' is selected. The main area displays 'SVM based on Terms and Concepts' for a document titled 'Delaware_scientists_create_shortest_ever_metal_to_metal_bond.txt'. The document text is shown with several terms highlighted in blue boxes. Below the text, a list of these highlighted terms is provided, including 'compound', 'collaboration', 'recently', 'preparation', 'long', 'connect', 'include', 'graduate', 'kevin', 'general', 'stronger', 'double', 'single', 'occur', 'triple', 'nature', 'complex', 'find', 'laboratory', 'make', 'react', 'structure', 'group', 'unusual', 'give', 'product', 'mix', 'weaken', 'reduce', 'length', 'previous', 'feature', 'source', 'bond', 'customs', 'warehouse', 'measuring equipment', 'report', 'Research Report', 'scientific profession', 'student', 'metals', 'plate', 'record', and 'recording'.

Explaining a text classification prediction made by PoolParty Semantic Suite, highlighting positive concepts and terms.

14

Structured data meets text

Knowledge Graphs as a Data Model for Machine Learning



15

USE CASES

Bringing Machine Learning, NLP and Knowledge Graphs together

16

The LinkedIn Economic Graph



The LinkedIn Economic Graph is a digital representation of the **global economy** based on

- ▶ 560 million members,
- ▶ 50 thousand skills,
- ▶ 20 million companies,
- ▶ 15 million open jobs, and
- ▶ 60 thousand schools.

<https://economicgraph.linkedin.com/>

“LinkedIn has a vast quantity of data. While much of the **data is structured**—graph nodes and edges, normalized fields in database records—a great deal of it is simply **natural language text**. Attaching structure and meaning to this text is essential to LinkedIn’s overall mission of connecting its members to opportunity.”

17

EventAdvisor

helps to 'putting out feelers' to find the right people and interesting content



ESCO

European Skills/Competences, qualifications and Occupations



13485 skills /
competences

2942
occupations

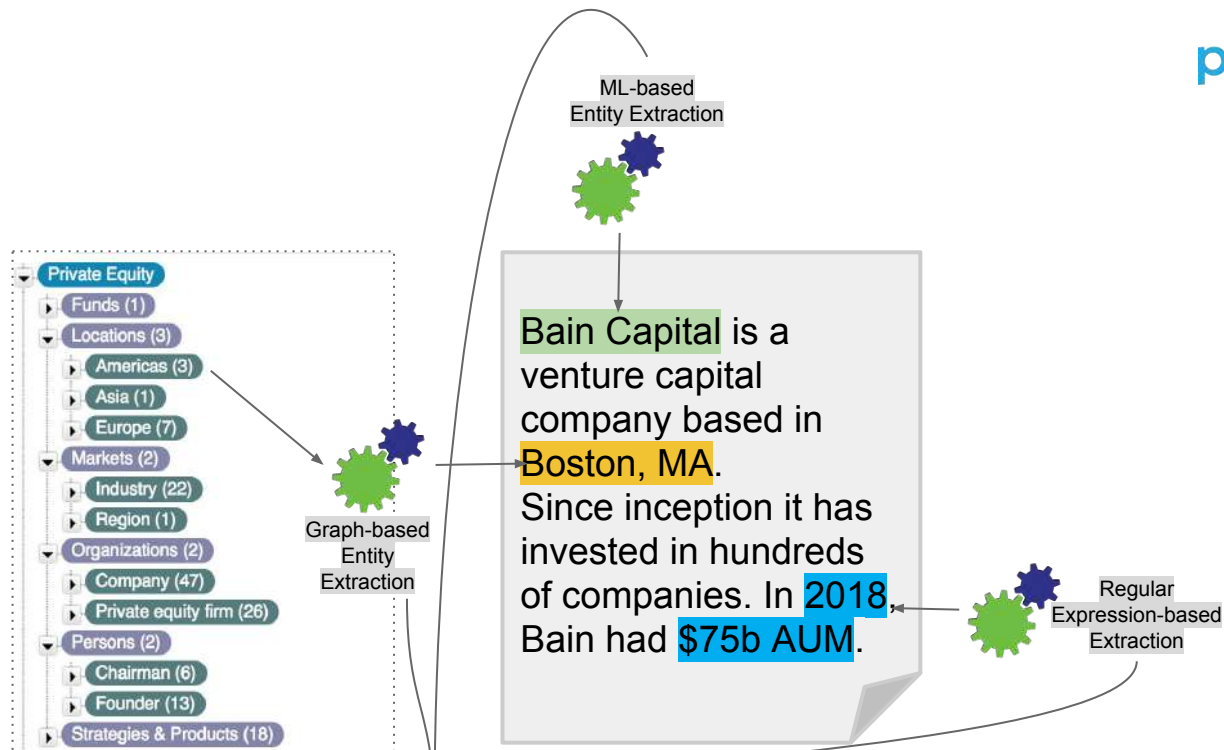
For more information
go to:

[https://ec.europa.eu/
esco/portal/home](https://ec.europa.eu/esco/portal/home)

18

Deep Text Analytics

Combining ML-based and Graph-based text mining with rules



SHACL-based Rules Engine

Give me all paragraphs in documents talking about American Private Equity firms with AUM higher than \$20b



19

Deep Text Analytics

Contract Validation



The screenshot shows the 'Contract Validation' interface. At the top, there are two main sections: 'Select files' and 'Choose rules'. The 'Select files' section contains a dashed box with the text 'Click to upload files you want to verify'. The 'Choose rules' section has a list of rules with checkboxes: Confidentiality (checked), Penalty (checked), Step-in rights (checked), Suspension right (unchecked), Termination for cause (unchecked), Termination for convenience (unchecked), and Partial Termination (checked). Below these are 'Run' and 'Clear' buttons. The 'Uploaded Files' section shows a single file: 'CompanyMasterAgreement.pdf - 4500216 bytes'. The 'Results' section is a table with three columns: file names, detected conditions, and their status. A tooltip is visible over the 'Matching part of contract' column.

File	Results	Status
CompanyMasterAgreement.pdf	Partial Termination	CONDITION NOT FOUND
CompanyMasterAgreement.pdf	Step-in rights	CONDITION NOT FULFILLED
CompanyMasterAgreement.pdf	Confidentiality Information	CONDITION FULFILLED
CompanyMasterAgreement.pdf	Confidentiality Period	CONDITION FULFILLED
CompanyMasterAgreement.pdf	Penalty	CONDITION FULFILLED
CompanyMasterAgreement.pdf	Confidentiality Clause	CONDITION FULFILLED
CompanyMasterAgreement.pdf	Step-in period	Matching part of contract: ED

Matching part of contract: ED
Confidentiality 23.1 The Parties agree to treat as confidence Confidential information. Both Parties

- ▶ Entities get extracted based on Machine Learning, Semantic Knowledge Models and/or regex
- ▶ Based on SHACL (W3C standards) conditions and rules are translated from SME into machine processable 'shapes'

21

Why Data Scientists need Semantic Models



- ▶ *Data Quality & Date Governance*
 - ▷ Content aboutness in a defined framework
 - ▷ Data relationships and context within a unified organizational model
 - ▷ Connections across disparate datasets
- ▶ *Improved Machine Learning*
 - ▷ Hierarchical or other mapped relationships allow for recommending similar content when exact matches not found
 - ▷ Granularity allows for more specific recommendations
 - ▷ Consistency across structure results more precise analysis and predictions

Source: [Suzanne Carroll](#), Data Science Product Director at XO Group

22

One question
at the end

Will Artificial
Intelligence
make
Subject Matter
Experts
obsolete?



Imagine you want to
build an application
that helps to identify
patients and
treatments pairings.

Which will you prefer?

1. Solely based on machine learning,
2. based on doctors' knowledge only,
3. or a combination of both?

23

Thank you for
your interest!



Andreas Blumauer

CEO, Semantic Web Company

- ▶ Mail andreas.blumauer@semantic-web.com
- ▶ Company <https://www.semantic-web.com>
- ▶ LinkedIn <https://www.linkedin.com/in/andreasblumauer>
- ▶ Twitter <https://twitter.com/semwebcompany>