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!
## 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

### 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.semantic-web.com**

**www.poolparty.biz**
#SemanticAI: Bringing Machine Learning, NLP and Knowledge Graphs together
AI suffers from a lack of common sense, but ...
... is talented in solving isolated problems based on isolated data sets

- Face recognition
- Deep Learning
- Genetic Algorithms
- Game AI
- Monte Carlo TS
- Deep Learning
- Neuronal networks
- Case based reasoning
- Fraud detection
- Credit cards
- Face recognition
- Game AI
- Fraud detection
“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.”
<table>
<thead>
<tr>
<th>Level</th>
<th>Example</th>
<th>Questions</th>
</tr>
</thead>
</table>
| (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 ...? |

Bloom’s Taxonomy: Classify cognitive processes
Perth is one of the most isolated major cities in the world, with a population of 2,022,044 living in Greater Perth.

Australia is a member of the OECD, United Nations, G20, ANZUS, and the World Trade Organisation.

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

“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
Six Core Aspects of Semantic AI

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.
Data Quality

Training data is semantically enriched with help from semantic knowledge models.

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

<table>
<thead>
<tr>
<th>Features used</th>
<th>Classifier</th>
<th>F1 (5-fold)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms</td>
<td>LinearSVC</td>
<td>0.83175</td>
<td>0.0008</td>
</tr>
<tr>
<td>Concepts from REEGLE + Shadow Concepts</td>
<td>LinearSVC</td>
<td>0.84451</td>
<td>0.0011</td>
</tr>
<tr>
<td>Concepts from REEGLE</td>
<td>LinearSVC</td>
<td>0.84647</td>
<td>0.0009</td>
</tr>
<tr>
<td>Terms + Concepts from REEGLE + Shadow Concepts</td>
<td>LinearSVC</td>
<td><strong>0.87474</strong></td>
<td><strong>0.0009</strong></td>
</tr>
</tbody>
</table>
No black-box

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)

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)

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

Knowledge Graphs as a Data Model for Machine Learning

14
USE CASES

Bringing Machine Learning, NLP and Knowledge Graphs together
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.”
EventAdvisor helps to 'putting out feelers' to find the right people and interesting content.

13485 skills / competences

2942 occupations

For more information go to: https://ec.europa.eu/esco/portal/home
Bain Capital is a venture capital company based in Boston, MA. Since inception it has invested in hundreds of companies. In 2018, Bain had $75b AUM.

Give me all paragraphs in documents talking about American Private Equity firms with AUM higher than $20b.
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’
DALICC stands for Data Licenses Clearance Center. It is a software framework that supports legal experts in the legally secure reutilization of third party data sources. The DALICC framework supports the automated clearance of rights.
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
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?
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