



# From Taxonomies to Recommendation Systems

The background of the slide is a light grey-blue color. It features a top-down view of a person's hands working at a desk. One hand is on a tablet displaying a colorful circular chart with four segments (green, blue, red, yellow) and numbers. Another hand is on a tablet displaying a grid of small profile pictures. A network diagram with white nodes and lines is overlaid on the right side of the image. A large globe icon with a 'T' inside is also visible.

**Heather Hedden**  
Data & Knowledge Engineer  
Semantic Web Company

# Semantic Web Company (SWC) and PoolParty



**SWC** is developer / vendor of **PoolParty Semantic Suite**

Most complete and secure **Semantic Middleware / Semantic AI platform** on the Global Market

**W3C** standards compliant



**ISO 27001:2013** certified

**First release in 2009**

Current version **8.0**

**On-premises** or **cloud-based**



Over **200** installations world-wide



**Semantic AI:**

Fusion of Graphs, NLP, and Machine Learning



Named as Visionary in **Gartner's Magic Quadrant** for Metadata Management Systems 2019, 2020



**KMWorld** listed PoolParty as one of the **Trend-Setting Products** 2015 - 2020 and listed SWC in the **AI 50** list of companies in 2020

- ▶ Why Recommendation Systems
- ▶ HR Recommender Example Demo
- ▶ How a Recommendation System is Built: HR Recommender
  - ▶ Taxonomy and Ontology Development
  - ▶ Text Mining: Extractor Service
  - ▶ Knowledge Graph and Search Application
- ▶ Other Recommenders: Matchmaking Systems

## Getting the right information to the right people

- ▶ There is a lot of information and content people can benefit from.
- ▶ They don't know how best to look for information that would benefit them.
- ▶ They don't know that the information is there or how to find it.

## Making matches of what goes together

- ▶ Standard search does not support complex matching queries.



A system that provides **suggestions** or **recommendations** to users can be very helpful.

# Why Recommendation Systems

---

**A recommender system (engine) can recommend to its users:**

- ▶ content of interest
- ▶ products to purchase
- ▶ people to connect with
- ▶ job opportunities
- ▶ training to improve skills



**A match-making kind of recommender system can recommend:**

- ▶ matches of applicants to job openings
- ▶ matches of consultants to projects
- ▶ matches of buyers and sellers

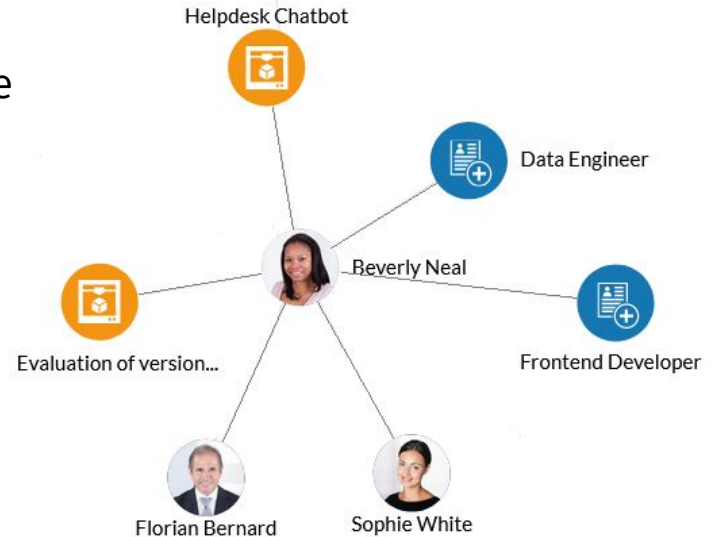
# Why Recommendation Systems

## HR Recommender System

A semantic matchmaking tool based on a knowledge graph

### Use case

- ▶ An organization wants to make the best use of the strengths and skills of its employees
- ▶ HR staff should be able to:
  - ▷ Find candidates for open positions
  - ▷ Staff projects
  - ▷ Identify professional development needs
- ▶ Employees should be able to:
  - ▷ Connect with interesting coworkers
  - ▷ Browse relevant projects
  - ▷ Find career opportunities within the organization



Demo

HR Recommender front end

<https://hr-recommender.poolparty.biz/>

## Recommender Technologies

- ▶ **Content-based filtering** - Similar content recommended based on a single user's interactions
  - ▶ Can only make recommendations on previous interactions or feedback of the user
- ▶ **Collaborative filtering** - Recommendations based on interactions from multiple similar users
  - ▶ Requires a large number of users

Disadvantages to both content-based and collaborative filtering:

- ▶ New users or items which had not been trained upon, don't get recommendations initially: "cold start" problem due to insufficient data
- ▶ Can only recommend to the user and not do other matchmaking

- ▶ **Support Vector Machines (SVM)** - Machine learning classification method, using algorithms, training examples, statistical learning, which calculates distances between categories (Often used in combination with **collaborative filtering**)
  - ▶ Requires time to train data, and performance varies based on the data.
  - ▶ Designed for limited, distinct classes; doesn't have the benefit a taxonomy with synonyms

- ▶ **Knowledge-based systems** - Based on explicit knowledge of the content, stored in a database



# How a Recommendation System is Built

---

## HR Recommender Components

### Semantic model

- ▶ Taxonomies containing concepts and labels ([PoolParty Thesaurus Management](#))
- ▶ Ontology of semantic relations ([PoolParty Ontology Management](#))

### Content that is text-mined

- ▶ CVs, personal profiles, job descriptions, project descriptions ([PoolParty Extractor](#))

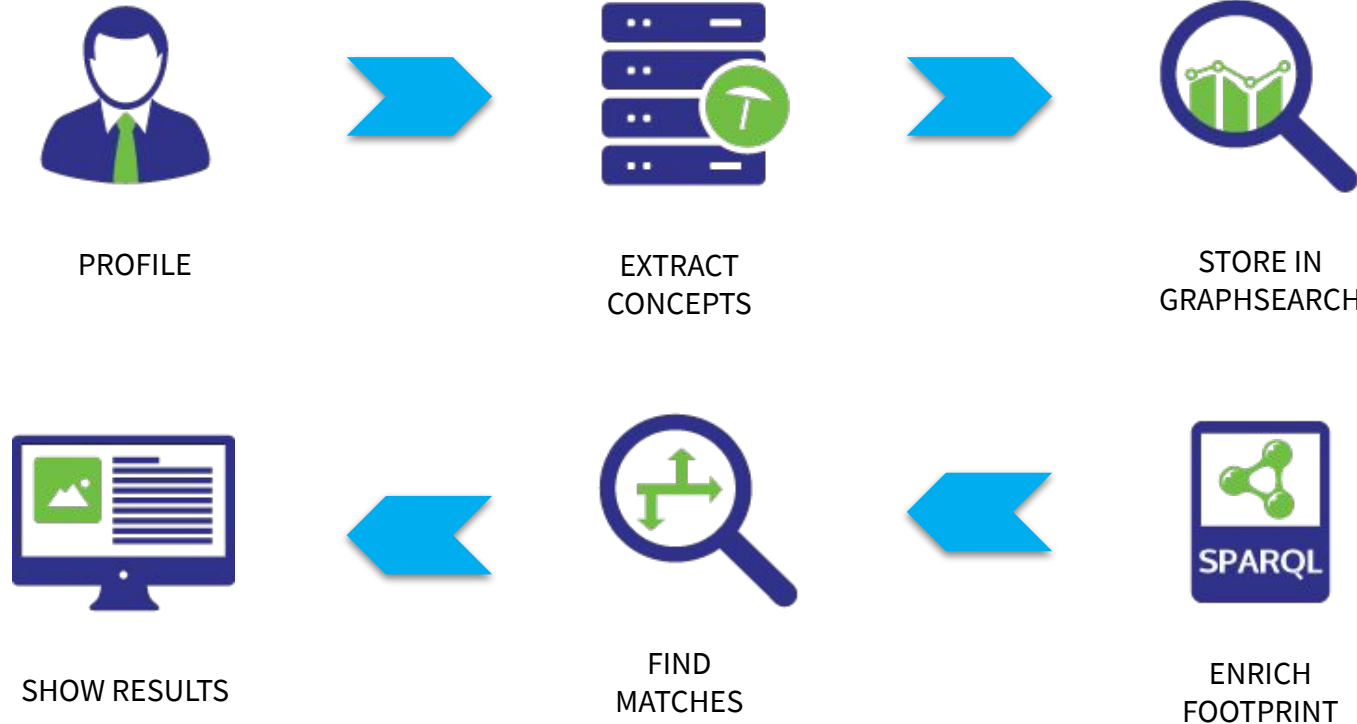
### Stored data

- ▶ Knowledge graph and a Solr search index ([PoolParty GraphSearch](#))

### Recommender application

- ▶ Algorithms for calculating similarities and recommendations to *enrich* the semantic footprint ([PoolParty SPARQL endpoint](#))
- ▶ Web application user interface on top of an API ([PoolParty GraphSearch API](#))

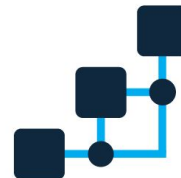
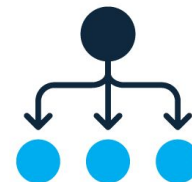
# How a Recommendation System is Built



## Taxonomy & Ontology for HR Recommender

### Taxonomy created from multiple sources

- ▶ Fully developed taxonomies
  - ▷ ESCO (<https://ec.europa.eu/esco>)
  - ▷ SEMWEB custom created taxonomy
- ▶ Enrich the taxonomy with text mining (entity extraction)
  - ▷ Propel
    - Industry conference content: submitted papers, speakers
    - Fictitious CVs



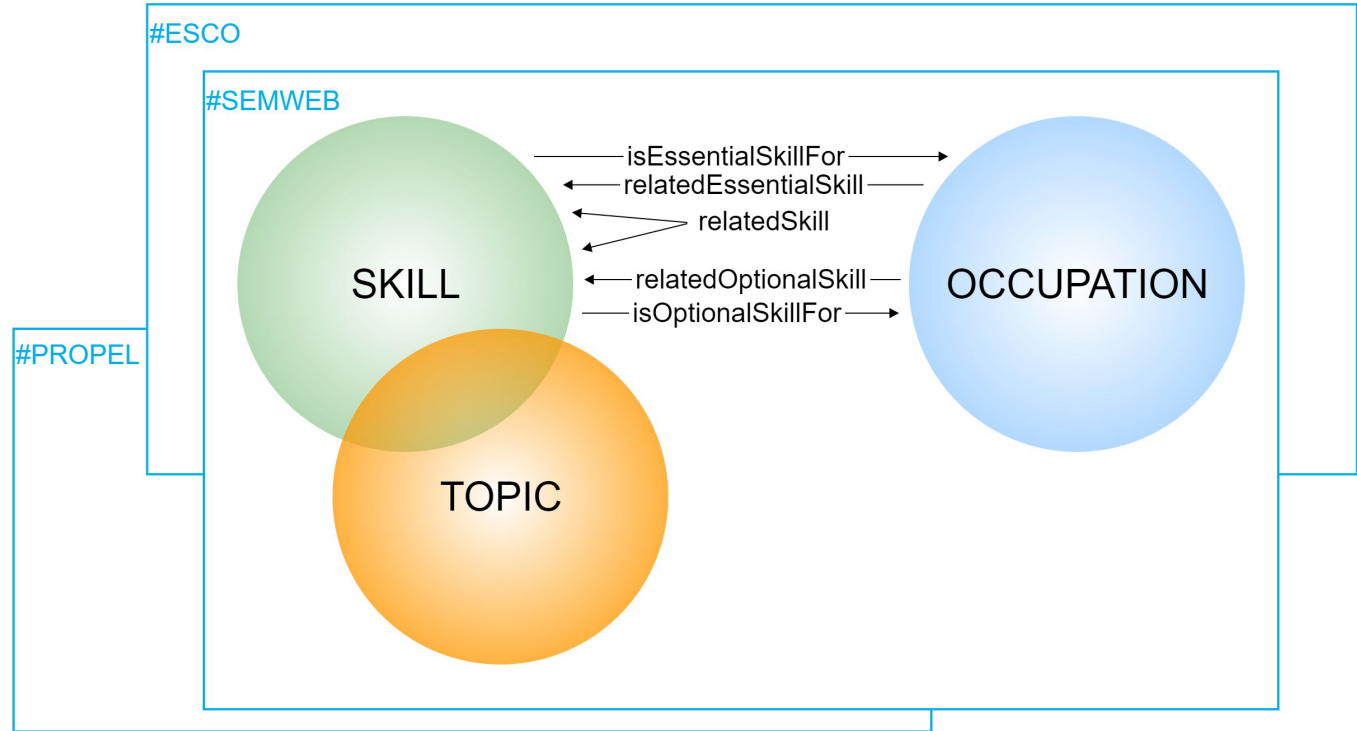
### Ontology layer to add semantic relationships

## Taxonomy sources:

- Skills & Occupations Topics: **SEMWEB** custom taxonomy
- Skills & Occupations: **ESCO** Classification
- Taxonomy enriched with text mining (term extraction) of Topics: **Propel** corpus of industry conference content: submitted papers, speakers

## Ontology layer:

Adds semantic relationships



Demo

HR Recommender back end taxonomy and ontology

in

PoolParty Thesaurus management

PoolParty Ontology management



## What is text mining?

- ▶ An application of text analytics, utilizing AI technologies of Natural Language Processing (NLP).
- ▶ Extracting passages from text that are relevant in a particular business context.
- ▶ Automatically deriving information, and not merely strings of words.
- ▶ Transforming unstructured text into meaningful information.

## PoolParty Extractor - text mining service

1. Extracts terms as candidate concepts from a corpus to enrich the taxonomy
2. Extracts taxonomy concepts for auto-tagging documents

## For HR Recommender

- ▶ Tagging documents of profiles, CVs, projects, and job openings

## Demo

### PoolParty Extractor

1. Corpus analysis extraction
2. Auto-tagging with Extractor API

# Stored Data in a Knowledge Graph

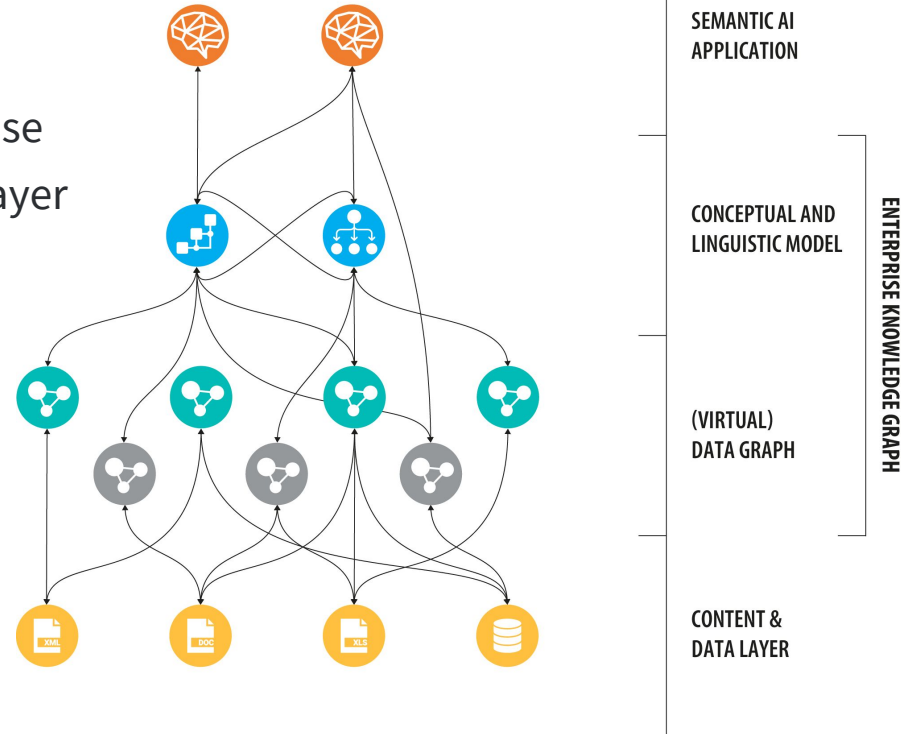
## What is a knowledge graph?

- ▶ Taxonomy + Ontology + Instance Data  
Often stored as triples in a graph database
- ▶ Connects to the content/external data layer and the semantic application layer

In the HR Recommender:

The semantic application is based on the Solr search Index.

Instance data are text snippets about each employee.





# Stored Data in a Knowledge Graph

## PoolParty GraphSearch

provides an API for search and recommender interfaces, including type-ahead, which is used for the HR Recommender.



PoolParty GraphSearch can also be used for displaying interactive search facets. (Not used in HR Recommender.)

# Application Build: Enrich the Footprint

## SPARQL query endpoint

Algorithms for calculating similarities and recommendations to *enrich* the semantic footprint

### SPARQL Endpoint

```
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX propel:<https://pp-semantic-dev.semantic-web.at/PROPELontology#>
PREFIX esco:<http://data.europa.eu/esco/model#>
SELECT *
WHERE {
  ?uri skos:prefLabel ?label .
  {
    SELECT ?uri (MAX(?distScore) AS ?maxDistScore)
    WHERE {
      VALUES ?x { <http://data.europa.eu/esco/skill/19a8293b-8e95-4de3-983f-77484079c389> }
      {
        BIND(?x AS ?uri)
        BIND(STRDT("1.00",xsd.float) AS ?distScore)
      } UNION {
        ?x esco:isEssentialSkillFor ?uri.
        BIND(STRDT("0.5",xsd.float) AS ?distScore)
      } UNION {
```

### Add Namespace

- SKOS
- DC
- DCTerms
- OWL
- RDF
- RDFS
- SWC

Run Query

```
uri
http://data.europa.eu/esco/skill/19a8293b-8e95-4de3-983f-77484079c389
http://data.europa.eu/esco/skill/eb0e5615-1575-4a86-a1a2-7d39595033c5
http://data.europa.eu/esco/skill/b4dc6e4f-dc7d-445f-8ce2-d7b9d225e282
http://data.europa.eu/esco/skill/58d7a289-dafd-4363-833f-d1dc4140885e
http://data.europa.eu/esco/skill/56a7f561-1d55-43c9-9cd7-36a0a9bc6c50
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/820a0683-d2ae-4a88-a648-3feb2f104e44
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/3bfcfe2-4f52-4b55-99d3-cb76a7e8131e
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/f31ae51f-aff9-42b1-81e9-fc9a55302090
http://data.europa.eu/esco/skill/47b9bbcf-356c-4782-83a4-7f5a1b2b51a3
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/3014fd67-33b2-4992-bffc-042338cdb026
http://data.europa.eu/esco/skill/4c016b68-4116-468c-9dc6-42710c239e4a
http://data.europa.eu/esco/skill/b633eb55-8f1f-4ae6-ab4c-2022fe2cb7f1
http://data.europa.eu/esco/skill/def007fa-5fed-4a5f-91a2-b0d7e3db1be1
http://data.europa.eu/esco/skill/993b1e23-f2de-4bd8-b33f-f86dde1c8e9d
http://data.europa.eu/esco/skill/0ccd6dcf1-5778-42a5-b685-4d01ae4a4871
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/661ac55d-3e7f-4cd8-bad3-4dc4af6efab0
https://pp-semantic-dev.semantic-web.at/PeopleandContentMatchmaker/8e7dabd3-bcd8-4309-adb0-dac16ee331db
```

label	maxDistScore
"Java"@en	"1.00"<<http://www.w3.org/2001/XMLSchema#float>
"ABAP"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"AJAX"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"APL"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"ASP.NET"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Angular"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"AngularJS"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Apex"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Assembly"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"C"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"C#"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"C++"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"COBOL"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"CoffeeScript"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Common Lisp"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Crystal"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>
"Delphi"@en	"0.7"<<http://www.w3.org/2001/XMLSchema#float>

## Matchmaking Demos

### Wine & Cheese Harmonizer

<http://vocabulary.semantic-web.at/GraphSearch>

### Semantic Matchmaker

(Matching consultants to projects)

<https://semantic-matchmaker.poolparty.biz>

Based on semantic data fabric

## Semantic recommender systems are based on:

- ▶ A knowledge graph comprising:
  - ▷ A taxonomy, whose concepts are tagged to and/or extracted from the content to be recommended *and* to either matchable content or a user profile
  - ▷ An ontology that links concepts with additional semantic relationships
  - ▷ Instance data linked to the taxonomy/ontology stored in a search index or graph DB
- ▶ A large body of content tagged with the taxonomy

## Optionally enhanced with:

- ▶ Algorithms for weighting/scoring relations

## And nice to have:

- ▶ A front-end (user interface) application



# Contact

---



**Heather Hedden**

[heather.hedden@semantic-web.com](mailto:heather.hedden@semantic-web.com)

[www.poolparty.biz](http://www.poolparty.biz)

[www.semantic-web.com](http://www.semantic-web.com)