

Cobrainer Technology Whitepaper

Hanns-Bertin Aderhold, Managing Director

Martin Kiechle, Head of Technology

The content herein is correct as of January 2020 and represents the status quo at the time it was written. Cobrainer's systems and any system metric to which reference is made herein will evolve going forward, as we continually improve functionality for our customers.

INTRODUCTION

The Cobrainer product is based on a proprietary, machine-learning-based skill profiling and skill matching technology. It enables large and medium size organizations to generate skill profiles for their employees from existing internal data. These skill profiles are then utilized to enable job, training, and project recommendations for employees as well as for recruiters. To deliver powerful, real-time skill profiling and job matching at scale, we've built a suite of tightly integrated components. In this whitepaper, we will give a brief overview of the four core modules of our technology:

- The Skills Graph System
- The Concept Recognition Engine
- The Recommendation Engine
- The Skill Curation Back-Loop

By the numbers:

- Cobrainer has built and is growing the world's largest skill database with 550.000 skill concepts
- Each skill concept has an average of 14.5 label representations (=8 million skill labels in the database)
- Cobrainer currently stores 200 million weighted relationships between the skill concepts
- Any Cobrainer customer has a historic uptime for both ingestion and queries of 99.999%
- Cobrainer currently services an average of 50.000 requests to its system per day

550.000
Skills

>50.000
daily requests

SKILL GRAPH SYSTEM 1/2

Cobrainer's Skill Graph System is designed to enable a unique new approach to a system-of-record for available and non-available skills of an organization and it makes such skill records easily and instantaneously retrievable. The core of the Cobrainer Skill Graph System is the Universal Skill Graph which is a central, customer-independent, and general reference graph database built and managed by Cobrainer. The graph stores the following data categories:

- general reference skill concepts,
- general reference label representation of such skills, and
- general reference relationships between skill concepts.

These three data categories are continuously aggregated and extracted from public domain content sources, such as ArXiv, European Patent Office (EPO), Stack Exchange, Wikipedia and others. *Skill concepts* are the basic record of a skill; *label representations* persist different notations, languages representations and synonyms of a skill concept; *relationships* store contextual information (i.e. the 'Relatedness') between skill concepts.

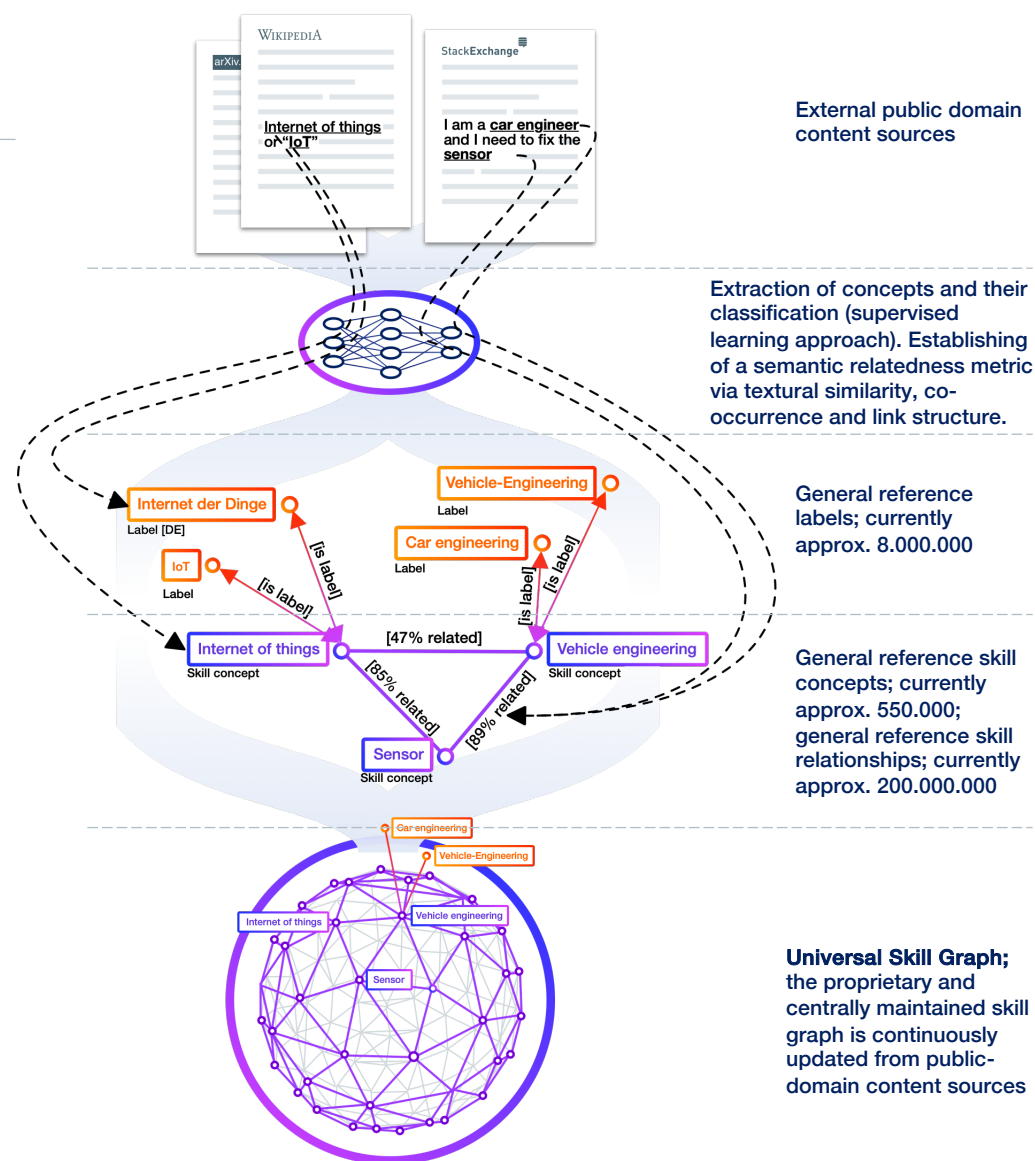


Figure 1: Universal Skill Graph Generation and Data Categories

SKILL GRAPH SYSTEM 2/2

Initially and upon start of an integration phase, every Cobrainer customer receives a dedicated and individual copy of the Universal Skill Graph as well as the artificial neural network for skill extraction. This customer instance of the skill graph system is maintained within a dedicated virtual server in the Cobrainer-operated cloud. A data link to the Universal Skill Graph is maintained, such that the customer skill graph continuously receives 1) *updates* to existing skill concepts, labels and, relationships, as well as 2) *new* concepts, labels, and relationships, and 3) *deletions* of certain concepts, labels, and relationships.

To ensure all relevant industry and customer skills (such as product names, acronyms, or unique process titles etc.) for an organization are represented in the customer skill graph, the system is fed with customer data (e.g. skill catalogues and job postings). This way, customer-specific skill concepts are detected and uniquely generated within the respective customer skill graph. Customers thus receive a skill concepts repository, that is continuously up-to-date from the 'outside world' and combined with customer-specific skills.

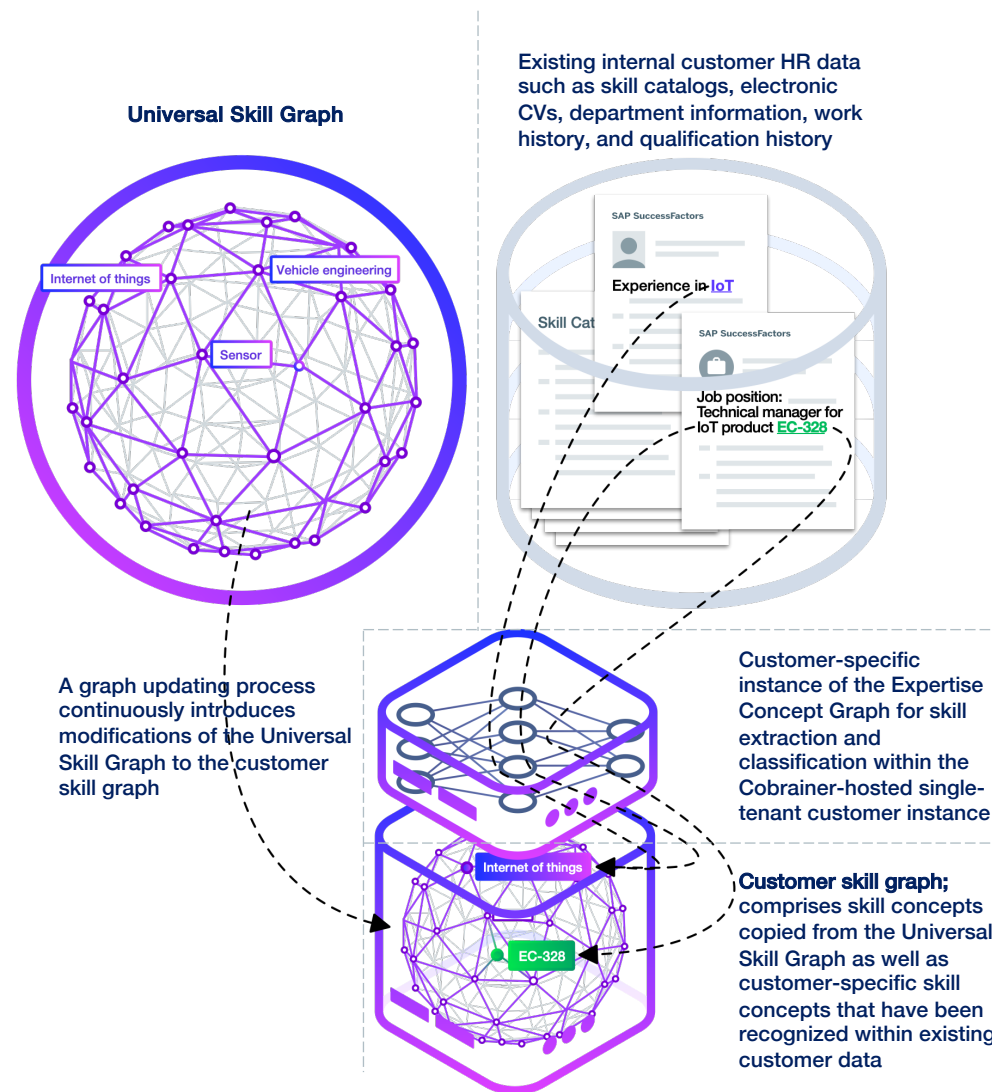
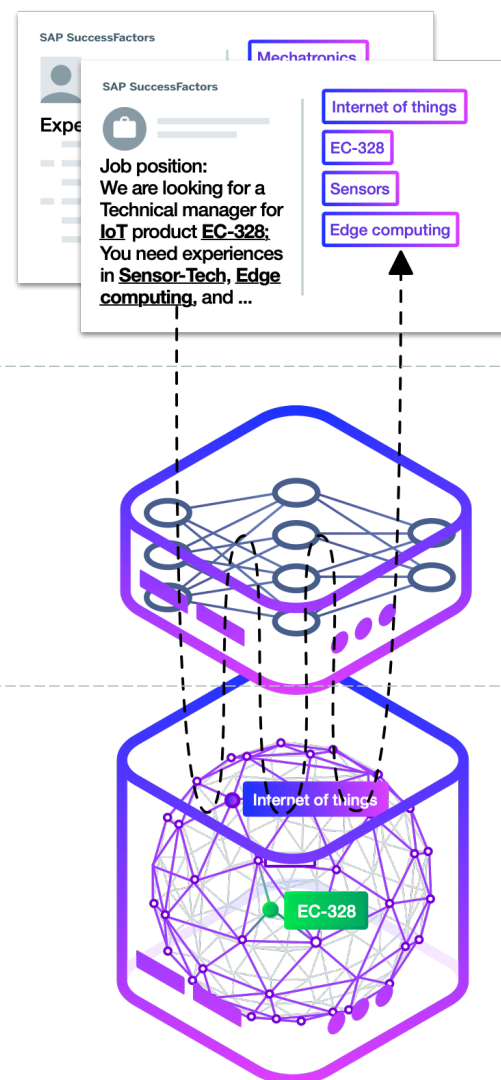


Figure 2: Customer skill graph data processing schema

CONCEPT RECOGNITION ENGINE

Once the customer skill graph system is set up, Cobrainer's Concept Recognition Engine extracts and generates automatic skill tag profiles from full text data for employees, job postings, project descriptions, or course descriptions. A unique characteristic of this process is that differences in wording or spelling of the full-text input data are offset and globally harmonized skill tags for the customer are generated. (E.g. The full-text input of the words "Car engineering" and "Vehicle engineering" are both converted to the harmonized skill concept "Vehicle engineering").

The Concept Recognition Engine helps recruiters to convert written job descriptions into actionable skill requirements within milliseconds. At the same time it helps employees to automatically generate a harmonized personal skill profile from their internal HR data, i.e. from SAP SuccessFactors, and –if desired– with external data, i.e. their LinkedIn or Xing profile.



Customer entities such as employee, jobs, projects receive unique and harmonized tag profiles from full-text input into the Cobrainer concept recognition engine

Concept recognition engine; an instance of the Artificial neural network for skill extraction and classification within the Cobrainer-hosted single-tenant customer instance

Customer skill graph which comprises a subset of the Universal Skill Graph and customer-specific skill concepts that have been recognized within existing customer data

Figure 3: Concept Recognition Engine data processing schema

RECOMMENDATION ENGINE

The Recommendation Engine is the underlying process for suggesting relevant jobs, projects, and training courses to an employee or vice versa to suggest relevant employees to an open vacancy. Let's follow the lifecycle of a recommendation process:

1. An 'input-entity' for which recommendations shall be produced (i.e. an employee, a job, a project, or a training course) is located within the graph through a skill concept set generated from the Concept Recognition Engine.
2. The first step of a recommendation query for the 'input-entity' identifies other 'output-entities' with skill concepts set in vicinity of the graph.
3. The second step of a recommendation query ranks all identified 'output-entities' in the vicinity to the 'input-entity' via path queries between the respective skill concept sets of the 'input entity' and the 'output entities'. Path queries represent the contextual 'nearness' between skill concept sets.

Employees, job positions, projects, and courses are 'located' on the customer skill graph by their attributed skill concepts; any recommendation is a type of graph query that identifies other entities 'near' to a given entity

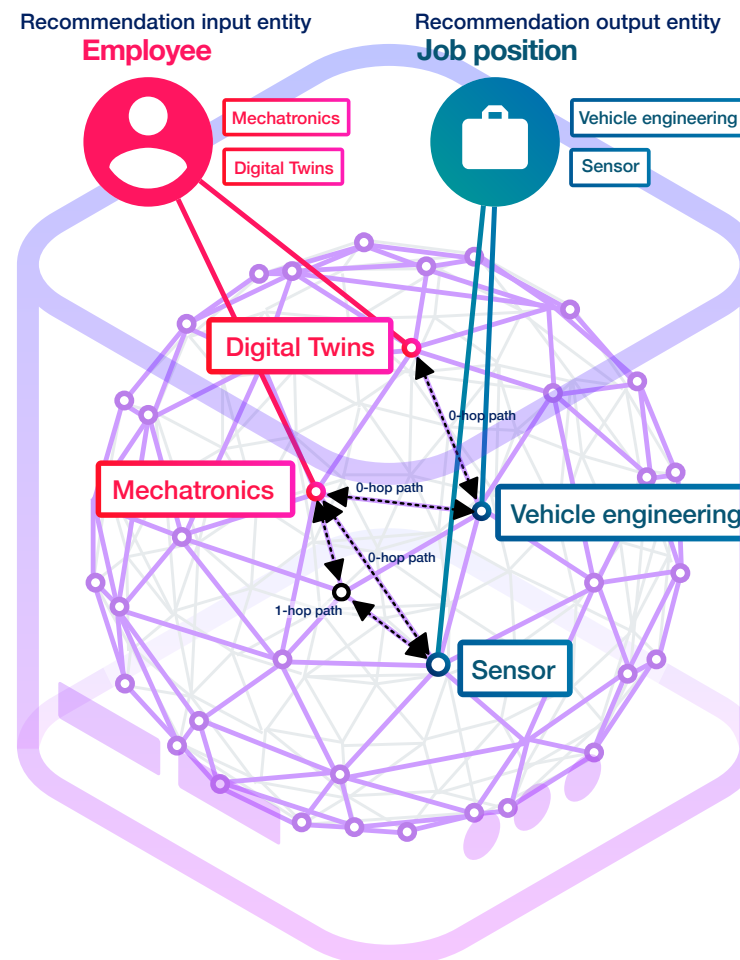


Figure 4: Recommendation Engine operating principle

SKILL CURATION BACK-LOOP

The Skill Graph System is set up as a dynamically improving system, meaning that certain user interactions are anonymously aggregated in the Universal Skill Graph System and utilized to improve the quality of the Concept Recognition Engine and the Recommendation Engine. This Skill Curation Back-Loop, is a continuous process set up between all customer instances and the Universal Skill Graph.

Customers are closely informed of such interaction data feeds and individually consent to such processes. The intrinsic incentive for consent is quality gains from the data network effect achieved by the aggregate of all Cobrainer customers.

Such data network effects, in which user interaction further improves the quality of the Skill Graph System, increase the market entry barrier for Cobrainer competitors and achieve a significant quality increase of the Cobrainer concept recognition engine.

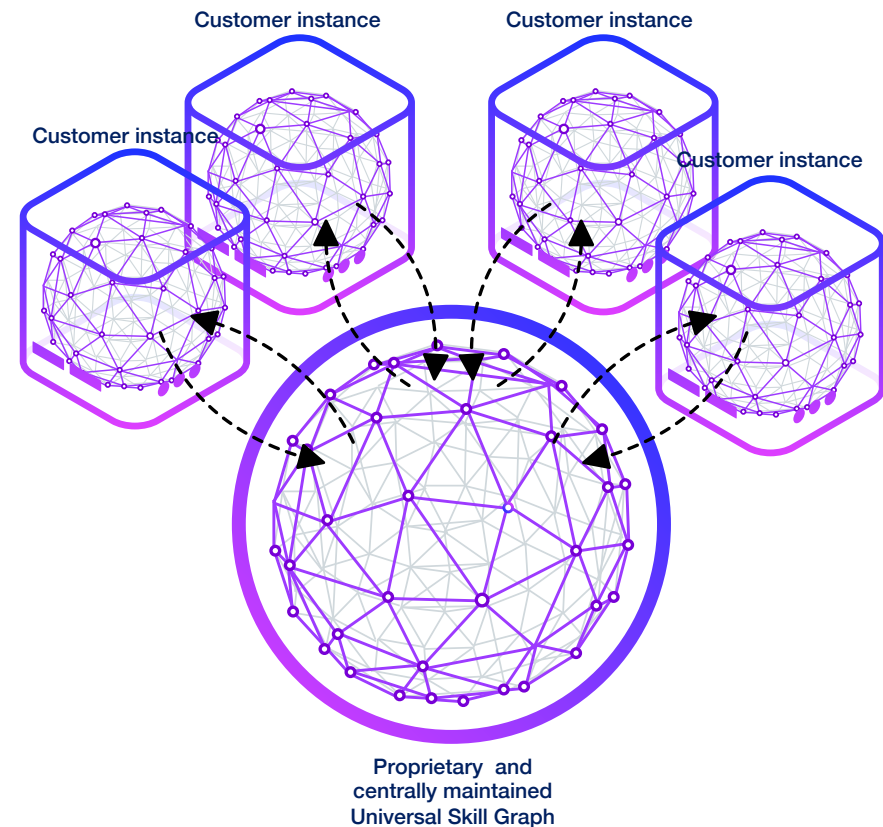


Figure 5: Skill Curation Back-Loop mechanism

SUMMARY

The Cobrainer technology is a tool that puts employees in the center of organizations by analyzing skills individually and providing skill-related transparency in the ever-more complex structure of large organizations. The Cobrainer technology suite achieves that with high accuracy and speed by providing to each customer a unique skill graph and a pre-trained tagging and recommendation engine that improves constantly via a well-structured data network effect.

