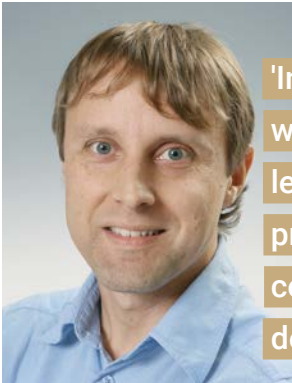


# Optimising data acquisition processes using big data

The DISCOVER project develops methods for automatic data acquisition, extraction and integration of decision-relevant information from heterogeneous online sources, which are also capable of analysing content from the deep web.



'In combination with artificial intelligence and deep web mining, big data can significantly increase the level of automation and efficiency of information procurement processes so that more up-to-date, comprehensive and high-quality data is available for decision-making processes.'

Prof. Dr. Albert Weichselbraun, Project Manager and lecturer at SII

## Project

DISCOVER – Knowledge discovery, extraction and fusion for improved decision making

## Lead

Swiss Institute for Information Science (SII)

## Project Manager

Prof. Dr. Albert Weichselbraun

## Team

Norman Süsstrunk, Philipp Kuntschik, Adrian Brasoveanu, Fabian Odoni

## Research Field

Data Analytics

## Commission/financing

Innosuisse, Venture Valuation VV AG

## Duration

December 2016 – January 2019

## Background

Venture Valuation AG provides independent evaluations of pharmaceutical and biotech start-ups and their products and technologies. The company also operates Biotechgate, a platform that offers comprehensive information on stakeholders from the biotechnology, pharmaceutical and medical technology sectors in an aggregated and structured form. This includes, for example, information on the product pipeline of companies, their financing, license agreements they have concluded or the contact information of their management team. Biotechgate's data volume has increased significantly in recent years, whereby existing data sets are also subject to constant change, which requires extensive investment in data acquisition and curation.

## Project objective

The DISCOVER project develops components that automate data procurement processes, thus significantly boosting their efficiency. The focus is on expanding Biotechgate to include information on clinical trials, improving data currency and quality and reducing costs for data acquisition.

## Implementation

Significant sections of the World Wide Web are not accessible to search engines, as the corresponding web resources are fed from extensive topic-specific databases. In practice, these data sources, also referred to as the 'deep web', are often particularly relevant because they usually contain extensive, high-quality and extremely specific specialist information. The DISCOVER project has developed methods for automatic data acquisition, extraction and curation, which are also capable of analysing information from the deep web. The system accesses domain-specific background knowledge, which is encoded in ontologies, databases or economic models, for example, so that searches for deep web resources can be optimised.

This will allow the DISCOVER pipeline to access information on clinical trials published on the WHO clinical trials platform. The sequencing of accesses is determined by domain-specific models. In the next step, knowledge extraction methods analyse the mirrored content to extract relevant information – such as study content, symptoms, and study progress – using text and data mining. The corresponding data records are then normalised and stored in Biotechgate.

further key DISCOVER component analyses the websites of all organisations available in Biotechgate, identifies the management and contact persons in these websites and compares the corresponding data sets with Biotechgate. This makes it possible to automatically detect changes in management and thus ensure that Biotechgate is up to date without this leading to higher costs for data curation. The websites are also searched for relevant publications on concluded licensing agreements, financing rounds or M&A activities so that this information can be made available to Biotechgate's customers in good time.

## Resultate

In der Praxis führen die im Rahmen des DISCOVER-Projekts entwickelten Innovationen dazu, dass den Kunden des Industriepartners aktuellere und umfangreichere Daten zur Bewertung von Biotech- und Pharmaunternehmen zur Verfügung stehen. Gleichzeitig konnte Biotechgate um klinische Studien erweitert und die Aktualität der Daten erhöht werden. Das DISCOVER-Projekt wurde durch Innosuisse gefördert. Diese Unterstützung hat massgeblich dazu beigetragen, Methoden der Grundlagen- und angewandten Forschung in kommerziell wertvolle Anwendungen eines innovativen Schweizer Unternehmens zu

«DISCOVER ist für uns ein entscheidender Schritt in der Digitalisierung der Datenbeschaffung. Dadurch kann unseren Kunden relevante Informationen in grösserem Umfang, noch schneller und zu niedrigeren Kosten zur Verfügung gestellt werden. Nicht zuletzt ist dies für uns ein wichtiger Wettbewerbsvorteil.»



Jost Renggli, COO und Mitinhaber Venture Valuation AG

## Schweizerisches Institut für Informationswissenschaft (SII)

Das SII beschäftigt sich mit Lösungen zu Fragestellungen und Problemen im Bereich der Produktion, Organisation und Distribution von Information und Wissen. Hierbei verfügt das interdisziplinäre Team des SII sowohl über das erforderliche Methodenwissen als auch über die notwendigen Kenntnisse aus verschiedenen Anwendungsdomänen in Wirtschaft und Verwaltung.

### Kontakt

Telefon +41 81 286 24 79  
E-Mail [sii@htwchur.ch](mailto:sii@htwchur.ch)  
Webseite [htwchur.ch/sii](http://htwchur.ch/sii)