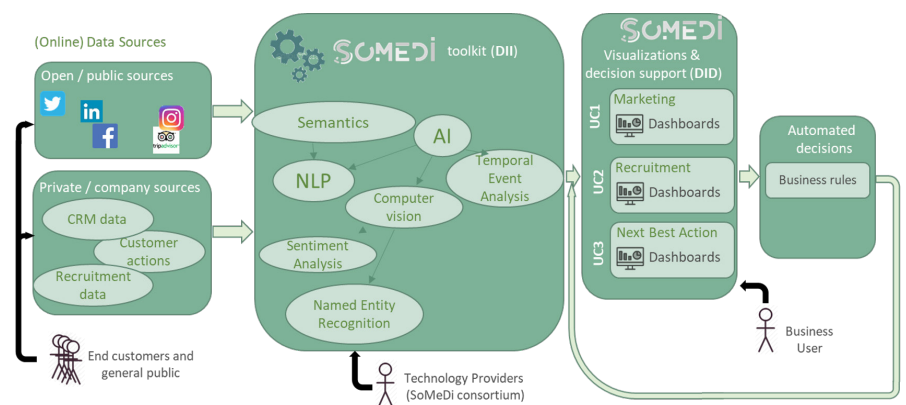# SoMeDi

Inferring and extracting hidden meaning

EXECUTIVE SUMMARY

To automatically serve customers with different preferences, it must be possible for companies to automatically understand their opinions. The ITEA project SoMeDi has therefore expanded Natural Language Processing in the form of a toolkit that understands the sentiment of human texts and images.

PROJECT ORIGINS

The digitalisation of business processes and private communication has led to huge amounts of digital interaction data: global IP traffic is expected to reach 4.8 ZB annually by 2020, or 396 exabytes (EB) per month. Although this offers companies new services in global markets, it has become difficult to find valuable insights in this mass of data. Digital Interaction Intelligence (DII), the extraction of information and hidden value through advanced data mining techniques, is therefore key to unlocking the full potential of digital interactions.

SoMeDi toolkit and its application in the three project use-cases

The SoMeDi (Social Media and Digital Interaction Intelligence) project is a toolkit that enables businesses to reach customers through digital means such as social networks, chatbots and mobile applications. By incorporating tools that can easily be combined and using techniques like Natural Language Processing (NLP), deep learning and computer vision to extract text and image data, the toolkit can be used to measure metrics such as brand popularity over time. The consortium offers these tools as services, ultimately delivering dashboards and recommendations to help them decide on their next best action.

TECHNOLOGY APPLIED

80% of textual data is currently unstructured. The centrepiece of SoMeDi has therefore been the development of NLP, a subset of Artificial Intelligence (AI) allowing machines to derive meaning from human languages. NLP was already well-developed for English and Spanish, so tools could be built upon open source developments. Turkish and Romanian NLP, however, were developed from scratch.

NLP has been utilised in three use-cases. In the first (Marketing), a restaurant chain wanted to examine how people communicate with them on social media and the impact of their advertising. SoMeDi developed a new AI and deep learning-based approach to Sentiment Analysis (SA), allowing machines to determine the attitude of a text towards a subject. SA was also applied in the Recruitment use-case, in which a SA microservice analyses a candidate's profile information and internship application and matches their skills to different internship campaigns. BEIA advanced a SA microservice, which offers the possibility to

use multiple SA engines and run text analytics for content in several languages and in different domains.

The final use-case, Next Best Action, helps telco companies to match their voice and data offerings with a customer's usage. This requires very large datasheets, so streaming processing and NLP techniques have been utilised. The platform gathers the real-time data in the form of customer behaviours on different channels and uses this to develop machine learning models and algorithms.

MAKING THE DIFFERENCE

SoMeDi provides businesses with a competitive advantage, as understanding users' intentions allows services to be tailored to their needs. HI-Iberia, for instance, has used SoMeDi to extend

their HIPPOS restaurant management software with social media analysis; this extra service is provided for free to give them a competitive edge. Turkcell Technology and Evam have developed a Real-Time Interaction Management Platform which has already replaced existing platforms in Turkcell, improving their business rule-generation time from four to six days (human-generated) to less than two hours (AI-enabled). It will be offered to telecommunications companies in Western Asia and North Africa.

Turkgen has created and commercialised its novel Social Media Data Analysis solution. This achieves highly accurate extraction with a more than 85% success rate by collecting data from many different sources, categorising this big data in near-real time and instantly storing it in NoSQL databases. Semantik developed a Telecom Chatbot trained for a unique telecom dataset. This decreases human interaction by between 10-20% compared to human answering agents.

BEIA and SIVECO are currently using a recruitment tool which has reduced the time taken to fill internships from up to two months to 25 days. Another key result of the recruitment application was reducing the time spent on validating new business rules from ten days to two and the time

to change configurations from two days to almost two hours. New analytical models can be deployed to production in an average of five days instead of two months, while SA accuracy (as an F1-score) has been increased from 0.5 to 0.66 compared to human-evaluated data.

There are also clear benefits for customers: Spanish restaurant chain LATERAL, a customer of HI Iberia, has reduced its time spent on social medial management by two hours per week using SoMeDi. The new NLP languages allow companies to enter the previously-inaccessible NLP market, which is expected to have a compound annual growth rate of 18.78% until 2023.

In regard to data security, SoMeDi enables companies to install data analytics technology locally and run it on European servers. This puts data in the hands of customers within a European legislative framework that compartmentalises risks. The project has now been added to Linked Data Models for Emotion and Sentiment Analysis (W3C CG), a community group for gathering tools, resources, vocabularies and use-cases for sentiment and emotion analysis, and is being considered for the creation of an upcoming recommendation. This will help to further proliferate this novel approach to data.

MAJOR PROJECT OUTCOMES

Dissemination
- More than 10 publications on the scientific results (ECAI 2018, eLSE 2019, INCERC 2019)
- Presence in more that 10 conferences and events (IoT Week Bilbao 2018, IMWorld 2017 and 2018, Smart City Expo World Congress 2018 and 2019)

Exploitation (so far)
New products:
- Social media monitoring platform for restaurants (tested at Spanish restaurant chain with more than 25 restaurants)
- Matchmaking system for recruitment based on sentiment analysis, in production in Romania (BEIA, SIVECO) for real-world internship programmes management
- Automatic rule based Next Best Action system for telcos (already in production in Turkey)

New services:
- Integration of Sentiment Analysis microservice to several call center solutions
- Content-based image descriptors integrated in NLP toolchain for social media analytics

Standardisation
- Participation in the W3G working group for Linked Data Models for Emotion and Sentiment Analysis. Usage of their standardised sentiment markup in the SoMeDi prototypes

Patents
- BEIA filed a patent application for the SA microservice titled 'Integrated platform for digital interactions through a computer system based on AI technologies for sentiment analysis'

Partners

Romania
Beia Consult International
Siveco Romania

Spain
HI Iberia
Innovati Servicios Technologicos
Taiger

Turkey
Evam
Semantik
Turkcell Teknoloji
Turkgen

Project start
December 2016

Project end
December 2019

Project leader
Raúl Santos De La Cámara, HI Iberia

Project email
rsantos@hi-iberia.es

Project website
http://somedi.hi-iberia.es/