EW-Shopp project: Supporting Event and Weather-based Data Analytics and Marketing along the Shopper Journey

Matteo Palmonari[1], Michele Ciavotta[1⊠], Flavio De Paoli[1], Aljaž Košmerlj[2],
and Nikolay Nikolov[3]

[1] University of Milan-Bicocca
{matteo.palmonari, michele.ciavotta, flavio.depaoli}@unimib.it
[2] Joef Stefan Institute
aljaz.kosmerlj@ijs.si
[3] SINTEF
nikolay.nikolov@sintef.no

Abstract. EW-Shopp is an innovation project, the aim of which is to build a platform for support of data linking, integration, and analytics in companies from the e-commerce, retail, and marketing industries. The project consortium joins several business partners from different sectors of e-commerce including marketing, price comparison, and both web and brick-and-mortar stores. The project is developing several pilot services to test the platform and inform its further development.

Keywords: Machine Learning · Data Integration · e-commerce · Visual Analytics.

1 The project

EW-Shopp[4] aims at providing support to companies operating in the fragmented European e-commerce ecosystem in order to connect, transform and integrate their data with external sources and use analytics to gain insights into their business. It is an innovation action project funded within the H2020 Research and Innovation program of the European Commission (*ICT-14-2016-2017, Big Data Public-Private Partnership: cross-sectorial and cross-lingual data integration and experimentation*). With its start in January 2017, it is now at the mid-point of its three-year duration.

The crux of the project is to foster small and medium-sized enterprises (SMEs), which represent 99% of all businesses in the EU, by building a platform which will deliver an end-to-end flexible solution to work with consumer and market data. In particular, to achieve novel customer and market insights, the platform will empower data manipulation, linking, and enrichment of business data with weather and event information, as a first enabler to create powerful

---

[4] http://www.ew-shopp.eu

analytical services. In fact, as a result of operating in a multi-lingual and multi-market environment, European companies (even those not operating through digital channels) commonly deal with large amounts of data, in multiple languages, acquired from different sources and sectors, or generated internally. Furthermore, their businesses are strongly impacted by external factors such as weather and global, as well as local events. To compete with international e-commerce giants, these companies must increasingly leverage this business data using modern analytics technologies to power and improve their services. Unfortunately, managing and integrating this heterogeneous data is prohibitively costly and time-consuming for a large number of companies, whose workforce generally does not possess the necessary skills. This is especially true for SMEs. Further, language barriers, lack of common models and shared systems of identifiers to interlink data, make these data integration tasks even more challenging.

The EW-Shopp platform attempts to simplify and streamline these tasks and level the playing field. We firmly believe this to be possible and give as an example the case of a project partner that, by using a predictive model built on top of integrated data about click-through rate of products, weather, and events, could design a service able to increase advertising of top-gear sport equipment in the days before a sunny weekend during the Tour De France.

2   The Consortium

The majority of the consortium consists of private companies from various sectors of e-commerce, covering the entire shopper journey from advertising and market research to sales and customer relations management for both online as well as brick-and-mortar stores. This includes: Ceneje Ltd. (SI), the manager of the largest price comparison shopping platform in Slovenia, Croatia, Serbia, and Bosnia and Herzegovina; BrowseTel Limited (UK), a provider of multi-channel communication services for customer relation management; GFK Eurisko (IT), the Italian branch of the fourth largest market research company in the world; Big Bang Ltd. (SI), the largest electronics retailer in Slovenia; Measurence (IR), a provider of sensor-based analytics solutions for physical locations; and JOT Internet Media (SP), a digital marketing solutions company.

Providing the technical and research expertise for the development of the platform the consortium includes: Engineering Ingegneria Informatica S.p.A (IT), the leading Italian software and services group; University of Milano-Bicocca (IT), focusing on semantic and interactive technologies for data linking; SINTEF (NO), the leading Norwegian research institute supporting data integration through their DataGraft[5] platform, and Jožef Stefan Institute (SI), the largest research institute in Slovenia overseeing all project analytics efforts and providing the event data source through the Event Registry[6] global media monitoring platform.

---

[5] https://datagraft.io
[6] http://eventregistry.org

3   Innovation

The data sources combined in the project offer numerous possibilities for innovative applications of knowledge discovery and data mining methodology. By linking and integrating the data sets of business partners and external data sources regarding weather and events, models can be constructed that power real-time responsive services for digital marketing, reporting-style services for market research, advanced data and resource management services for retail and e-commerce companies and their technology providers, as well as enhanced location intelligence services.

To guide the development of the platform and provide a basis for its evaluation, several pilot services are in development:

– **Pilot I - Enrichment of purchase information for web platforms:** By building a predictive model of user interaction on an online shopping portal in relation to external weather and event factors, we will enable the portal to run a reactive sense-of-urgency information service. For example, before a heatwave, we inform consumers that air conditioning sales commonly spike in such conditions and delivery could be delayed significantly.
– **Pilot II - Integrated platform for category and marketing optimization:** Combining data from a price comparison platform and a retailer will enable analysis and modeling of business actions such as marketing campaigns and discounts. This analysis will power a business-to-business service that will allow a retailer to use the wider market view of the price comparison platform to inform its category and marketing management.
– **Pilot III - External data access API and decision-making systems supporting customized campaigns:** Using weather and event-based predictive models for predicting customer response rate in a call center and managing marketing campaigns.
– **Pilot IV - Location Intelligence:** Modeling seasonal dependencies of visits to physical store locations to support activity planning and management.
– **Pilot V - Campaign-driven purchasing intentions:** Modeling the dynamics of web search engine keywords (e.g. Google Adwords[7]) with respect to weather and event factors to support marketing campaign management.

All the pilot services will be tested by business partners in the scope of their regular operations. Their feedback will guide development in the second half of the project.

As it is still undergoing development, we are unfortunately not able to show a full end-to-end interaction with the platform. Currently, we can demonstrate a selection of analytics results from the pilots and show the business insights they offer. These results will be presented with a strong focus on interactive visualizations, as one of the guiding principles of the project is an emphasis on intuitive and interactive visualizations in platform reporting services. Because they contain sensitive business information, we are unable to share these visualizations publicly before the conference.

---

[7] http://adwords.google.com

4   Platform

In this section, we present the characteristic elements of the EW-Shopp platform, for the design of which the consortium members have completed a requirements collection phase based mainly on the needs of the business partners (through the work done on the pilots) and on the best practices of the Big Data architectures.

At the end of this process, the need for an *open source* platform, capable of managing data in *tabular format* and of generating *linked data* to be used for analytics and visualization, became clear. In addition, two possible use cases have emerged. In the first case, small amounts of data have to be manipulated, the platform has to be lean and *easy to install on a commodity machine* similarly to tools like OpenRefine[8]. In the second scenario, the platform needs to manage big data, which means that a domain expert user must be able to describe the transformations through a user-friendly and interactive interface and to execute it in an automated way as a *batch* process (as in Karma[9] and Trifacta Wrangler[10]).

In an attempt to give a syncretic response to the needs outlined in the requirements, the platform has been conceived as composed of three macro-components (Figure 1):

Core Data Services (in light blue): These components provide access to corporate or third party data to be used in both data linking and extension processes. Figure 1 shows services to access the project's core data, i.e., weather (W), events (E) and products (P). Other enrichment sources include Wikifier[11], as well as freely accessible knowledge bases such as DBpedia[12].

Platform Services (in green): data preparation, analytics and visualization services. These services offer simple and intuitive user interfaces for creating (and executing if the working table is small enough) data wrangling, linking and extension pipelines.

Corporate Services (in red): These services implement platform components needed for data governance, i.e., ingestion, storage, processing, data flow and security management of massive data sets.

The platform components at application level in Figure 1 are:

Data Wrangler that is the component that enables the user to define the transformations of data cleaning, linking and enrichment at design time, and possibly on a data sample. Such data preparation processes will then be carried out by the component referred to as Big Data Runtime on the full dataset.

Data Analyzer that is the component that provides a set of predefined tools for predictive and prescriptive algorithms on enriched data.

---

[8] http://openrefine.org/
[9] http://usc-isi-i2.github.io/karma/
[10] https://www.trifacta.com/products/wrangler/
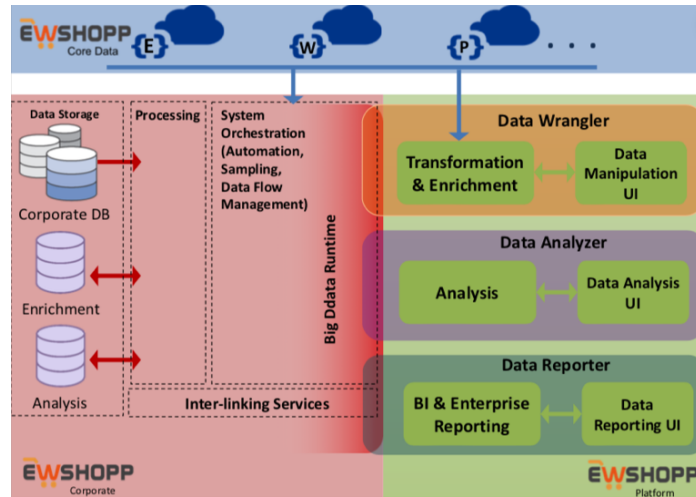[11] http://wikifier.org/
[12] https://wiki.dbpedia.org/

Fig. 1. General architecture

Data Reporter that is the component that allows the user to visualize and analyze the outcomes produced by the Analyzer from a business viewpoint.

With regards to the so-called Corporate services, these are gathered in a single macro-component called Big Data Runtime (BDR) that has the capability to execute transformation operations defined using the Platform services on genuine Big Data. In addition, it provides the data reporter with specific APIs for data access. Within the Big Data Runtime, we recognize a sub-component in charge of Data Storage and another dedicated to the Processing of such data.

5   Conclusions

The EW-Shopp ecosystem aims at fostering e-commerce, Retail and Marketing industries in improving their efficiency and competitiveness through providing a platform for supporting data linking, integration, and analytics in European e-commerce companies. It is developed in cooperation with businesses from all stages of a shopper's journey and will be tested in several pilot services to ensure that it addresses actual business needs. The project showcase will focus on selected analytics results collected during pilots' development and present them using rich interactive visualizations.

Acknowledgements