

HAVAS 18 Labs: A Knowledge Graph for Innovation in the Media Industry

José Gutiérrez-Cuéllar¹ and José Manuel Gómez-Pérez²

¹ HAVAS Media Group, Paseo de la Castellana 259 C planta 30
28046 Madrid, Spain

jose.gutierrez@havasmg.com

² iSOCO – Intelligent Software Components S.A., Av. Del Partenón planta 1, oficina 1.3A
Campo de las Naciones 28042 Madrid, Spain

jmgomez@isoco.com

Abstract. 18 months may seem a short time in absolute terms but in the corporate world and especially when referring to entrepreneurship 18 months is usually a figure associated to life expectancy, with 80% technology startups crashing and burning in their first 18 months. HAVAS has launched 18 Innovation Labs, a global initiative aiming to identify startups in the intersection of technology and media, in order to co-create new ways to revolutionize the media and entertainment industry. HAVAS seeks to interconnect startups, innovators, technology trends, other companies, and universities worldwide in a Knowledge Graph that supports analytics and strategic decision-making for the incorporation of such talent within their 18 months life span. In this talk we describe the 18 Labs initiative, challenges, and business expectations and how semantic technologies are key for realizing this vision by extracting startup information from online sources, structuring and enriching it into an actionable, self-sustainable semantic dataset, and providing media businesses with strategic knowledge about the most trending innovations.

Keywords: Media, startups, information extraction, aggregation, enrichment, linked data, SPARQL.

1 Overview of the talk

The communication between brands and consumers is set to explode. Product features are no longer the key to sales. Consumers want more. They want products to provide collective and personal benefits. As a matter of fact, those brands providing such value are more meaningful and derive clear economic benefits from being perceived as such. On the other hand, millennials, the digitally-savvy generation born between 1982 and 1999, are taking over, with a dramatic impact on how consumers and brands engage and what channels and technologies are required to enable this process. Millennials are by far the largest digitally native generation and rule over social media. As a result traditional boundaries within the media industry are being stretched and new ideas, inventions, and technologies are needed to keep up with the challenges raised by the increasing demands of this data-intensive, in-time, personalized, and thriving market.

It is therefore necessary to leverage advances in the area by stimulating a collaboration ecosystem between the different players. Inspiring examples include the adoption by Tesla Motors of an open patents policy, whereby Tesla shares their innovation in regards to electric cars openly via the internet¹. In return, Tesla expects the industry to use their learning and their intellectual property to further evolve the electric car industry and dynamize the market. In the media industry a clear example of this 'better together' approach is HAVAS 18 Innovation Labs, deployed at strategic locations around the world. One of such locations is the Siliwood² research center in Santa Monica, co-created in partnership with Orange, which focuses on the convergence between technology, data science, content and media. 18 Innovation Labs seeks to connect a great mix of local talent over the sites, involving innovators, universities, start-ups and technology trends to co-create initiatives relevant now and in the mid-term for both HAVAS and their clients to stay one step ahead.

With the help of iSOCO, their partner in semantic technologies, HAVAS is creating a knowledge graph and information platform that aggregates all the available knowledge about technology startups worldwide and makes it available for exploitation in a single entry point. We extract information from online sources, including generalist and specialized web sites, forums and blogs, online news, entrepreneurial and general purpose social networks, search engines and other content providers; structure and aggregate this information in an RDF dataset and enrich it by interlinking with external datasets; and provide an API for the exploitation of this knowledge by media business strategists in analytics platforms. Beyond factual knowledge about the different entities, the resulting knowledge graph makes emphasis on how such entities are related to each other. The relationships between them are described explicitly, supporting the discovery of new insights by navigating the graph.

In addition to automated information extraction means, the knowledge graph can also be populated with on-site information by local rapporteurs, usually members of the local entrepreneurial scene distributed at each of the HAVAS 18 Innovation Labs. Rapporteurs are provided with the means to introduce or modify new entities in the knowledge graph and define relations between them, according to the underlying data model. Rapporteurs are assisted by autocomplete functionalities based on the knowledge previously stored in the knowledge graph. Rapporteurs also play the critical role of curators of the knowledge graph information produced either by other peer rapporteurs or extracted automatically from online sources.

At the moment of writing this abstract, the HAVAS 18 Innovation Labs Knowledge Graph contains information about 1.812 startups, 559 technology trends, 1.597 innovators, 20 companies and 35 universities and research centers in the Siliwood area, following the Linked Data principles. All these entities are additionally connected to relevant online news, where they are mentioned (currently, 36.802), for extended and up-to-date information about them. The Knowledge Graph is updated daily in an automated batch process, identifying new entities and updating existing ones. We expect the knowledge graph to quickly reach the threshold of 300.000 startups below 18 months and extend to the remaining Labs in the next few months.

¹ <http://tinyurl.com/lsok7ew>

² <http://www.havasmedia.com/our-thoughts/blog/data/hooray-for-siliwood>