

Linked Data and Smart Recommendations for the Promotion of Independent Cultural Content

(In-use contribution)

Pierre Maret, Fabrice Muhlenbach (Université Lyon-Saint Etienne)
Cédric Claquin, Pierre-René Lhérisson (1D Lab Company)

Summary

Internet streaming has become the most popular media for consuming cultural content, but due to the reduced size and means of independent labels, users hardly ever access their content. So, we have started a Web of Data initiative targeting independent artists and labels. We aim to federate their contents and to run smart algorithms that help users discovering the novelties and enlarging their cultural appetite. The Linked Data approach perfectly suits the distributed creation of contents and necessity for labels to create links between their content (to get it all discovered). We take advantage of the knowledge representation format (RDF) and of reasoning capabilities to calculate various similarity measures and promote discovery.

Independent cultural content

The 1D Lab Company is specialized in the streaming of independent cultural content such as music and video (and soon videogames, comics and digital books). It explores, through a cooperative approach, new models of broadcasting and remuneration for indie content. The term "independent artist" refers, in the case of the music business, to an artist who is not produced by one of the 3 major companies. Independent artists are generally produced by independent labels. The number of labels in Europe is close to 20 000. They own approximatively 25 percent of the music produced (while 75% belongs to the "3 Majors").

Since streaming has become a natural way of accessing this content, the question of their indexing and discovery on the web is crucial, especially for independent artists and labels because they have no means for competing on the market against the Majors and for communicating with the global Internet companies. Independent labels and artists produce data, videos and diverse content in order to be discovered on the Web. But their obvious limited size and power, combined with the page ranked algorithm strategy, make it difficult for users to access this content. Essentially, artists and labels create novelty but they suffer from their small size and remain mostly invisible. Their data is not sufficiently connected to the rest of the Web.

Semantic web opportunities

Independent labels and artists (and structures promoting or consuming their content), should take advantage of a world wide data and knowledge base with several characteristics:

- The data and knowledge base should be initiated and moderated by professionals, it should be open;
- It should be able to receive and integrate the labels' catalogues;
- It should be enriched automatically with data gathered from the Web;
- Links should be established within the base and with data external to the base.

The Web of data techniques (Linked data, RDF/OWL, SPARQL) are well suited for the implementation of this base because they offer several characteristics which fit the needs well:

- A domain can be described as an ontology (types of data and types of links). Some ontologies already exist in the cultural domain, especially musicontology (musicontology.com). This ontology can be reused and progressively extended for further evolutions and needs. Also,

ontologies can be combined together, for instance muscicontology can be combined with Geospatial ontology to describe the locations of cultural events.

- The ontology is populated by actual data from the cultural domain (instances of the types of data and links). Datasets are composed of data/link types and instances. Datasets can be created independently.
- Each type and each piece of data can be published with a unique identifier (web URI).
- The modeling technique (RDF) is suited for establishing links between data within a dataset and between datasets (through web URIs to refer to the same objects). A network of data results from these links and automated reasoning capabilities can be used on it. An example of inferred data is the following: Artist A1 in dataset D1 is the same one as Artist A2 in dataset D2, and A2 plays American Folk music (dataset D2) which is influenced by Irish traditional music (Dataset D3).

Smart recommendations

The creation of an ontology and datasets for independent labels and artists only creates the opportunity to promote their content if it is associated with the implementation of smart algorithms. A crowd-based recommendation strategy may not be the right approach because crowd-sourcing naturally promotes popular content, which is never the case for independent content. Similarity-based recommendation is one of the key strategies for the promotion of independent content because it can lead to the discovery of all kinds of content and artists, known or unknown, recent or old. A linked data format is convenient for calculating complex similarities between artists or content descriptions thanks to the links established between data and thanks to reasoning capabilities. Distance functions involving natural language processing, text mining and data mining techniques can be established to this aim. Content analysis such as spectral analysis for sound or image analysis can also be introduced in these functions.

We have made a further step forward towards developing the discovery of cultural content by users: our aim is to help and incite (either directly or indirectly) the users to enlarge their cultural field or cultural appetite. This is a challenging issue because the proposed content should be on the one hand close enough to the user's interest, and on the other hand be different enough to give him/her the chance to discover the new fields he/she would not have discovered by his/her own. Specific distance functions exploiting links within web data have been developed to this aim.

Enterprise positioning and future works

The "new models" in streaming services are built at the expense of the vast majority of creators¹ and neglect notions such as open cultural novelty, artistic emergence and the development of territories. These notions are crucial for maintaining diversity. 1D Lab focuses only on independent contents so as to put forward such a wealth and get away from the logic of the mass market which makes it necessary for artists to accumulate millions of listenings in order to ensure a decent payment. 1D touch, the first fair streaming platform developed since 2013, aims at focusing on editorialized content to promote independent cultures. The data and knowledge base combined with the smart algorithms obviously participate to the development of this original business model. We plan to develop different tools for the constitution and the exploration of the data and knowledge base. We will also extend our recommendations to the promotion of local events (i.e. concerts, festivals) and to social exchanges (i.e. exchange of samples). We will combine different ontologies and datasets (such as FOAF, SIOC, Geospatial ontology) in order to introduce these aspects into our recommendation algorithms.

¹ only 1 % of the artists attract 77 % of the revenues generated by streaming (The death of the long tail - Midia consulting)
<https://www.musicindustryblog.wordpress.com>. The latest example has been given by Portishead with 34 million of streams for 2 000€.

Contributors

Pierre Maret is Professor in computer science at the University of Lyon-Saint Etienne. His fields of interests are Web based communities, Semantic web, and Recommendation systems.

Fabrice Muhlenbach is Associate Professor in Computer science at the University of Lyon-Saint Etienne. He is a specialist in data mining, recommendation systems and psychology.

Cédric Claquin is Deputy General Director of the start-up 1D lab, in charge of academic and research partnerships. An early activist of open culture and independent creations, former consultant in communication strategies, and manager of artists, he's involved more specifically in the research of new models integrating cooperative work and innovation.

Pierre-René Lhérisson is a researcher in computer science in 1D lab. He combines data mining, text mining and semantic web techniques.