

Researchers perceptions of DH trends and topics in the English and Spanish-speaking community. DayofDH data as a case study.

María Gimena del Rio Riande, Elena González Blanco García, Salvador Ros y Antonio Robles.

Cita:

María Gimena del Rio Riande, Elena González Blanco García, Salvador Ros y Antonio Robles (Julio, 2016). *Researchers perceptions of DH trends and topics in the English and Spanish-speaking community. DayofDH data as a case study. Digital Humanities 2016: Conference. Jagiellonian University & Pedagogical University, Cracovia.*

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/5>

ARK: <https://n2t.net/ark:/13683/pdea/mhf>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.



Digital Humanities 2016

Kraków, 11–16 July

(<http://dh2016.adho.org>)

DH Home(<http://www.dh2016.adho.org>) / Abstracts (</abstracts/>) / 358 (</abstracts/358>)

Show info

How to cite

XML Version (</static/data/230.xml>)

Robles-Gómez, A., González-Blanco, E., Ros, S., Del Rio Riande, G., Hernández, R., Tobarra, L., Caminero, A., Pastor, R. (2016). Researchers' perceptions of DH trends and topics in the English and Spanish-speaking community DayofDH data as a case study. In *Digital Humanities 2016: Conference Abstracts* Jagiellonian University & Pedagogical University Kraków, pp. 658-660.

Researchers' perceptions of DH trends and topics in the English and Spanish-speaking community. DayofDH data as a case study.

Defining the "state of the art" in Digital Humanities (DH) is a really challenging task, given the range of contents that this tag covers. One of the most successful efforts in this sense has been the international blogging event known as "DayofDH" or "A Day in the Life of the Digital Humanities" project, promoted and sponsored by centerNet (<http://www.dhcenter.net.org/>), which has put together digital humanists from around the world to document once a year what they do (Rockwell et al., 2012). The websites of DayofDH were hosted in North America until 2015, when it was coordinated in Europe by LINHD (<http://linhd.uned.es>), the Digital Innovation Lab, at UNED in Madrid. Participants belong to several countries around the world.

The relevance of DH in non-English speaking countries has been quick and important in the last decade, and especially important in the Spanish-speaking world (Spence and González-Blanco, 2014; González-Blanco, 2013; Del Rio Riande, 2014a; Del Rio Riande, 2014b; Galina et al., 2015). Technological projects for humanities have existed in the Spanish world for many years; however the discipline called "Digital Humanities" arose in 2011 with the first meeting that originated the Spanish Digital Humanities Association, HDH. This relevance is reflected in the creation of a parallel version of the DayofDH in Spanish, the "DíaHD", which was hosted by the UNAM in Mexico in 2013 and 2014 and converged in the last initiative at UNED transforming both blogging events into a bilingual version of the Day

Although there have been general studies about the information on participation in those events (Priani et al., 2014), there has not been an automated data analysis using NLP (Natural Language Processing) or Big Data tools to extract and classify the relevant information gathered in blogs (Tobarra et al., 2004). More technical details about these aspects can be found in (Tobarra et al., 2014b).

According to this, the main goal of this paper is to develop a dashboard that allows us to get more insight about interest topics and leaderships of this community during the period of time in which this event has been developed. With the "dashboard" word, we mean the analysis and presentation of results, not a tool. In this sense, the topic characterization process deals with the detection of the most relevant topics which are employed in the publication tools of these kinds of virtual communities (Tobarra et al., 2014a).

In order to achieve our aforementioned objectives, this work is focused on the datasets corresponding to four years of DayofDH (2012, 2013, 2014 and 2015 editions), and the Spanish version of the event in DíaHD 2013. This work strives at showing the evolution and trends in the last four years in order to give account of the presence of the Hispanic communities in the field. The information of the Spanish 2014 edition has been discarded, as it is not any more available online due to technical problems at the organizing institution. All editions of DayofDH employ WordPress, which has an associated SQL database, including several general tables and a specific set of tables per blog, defined in the project and common to all editions. The CMS is combined with the BuddyPress social plugin, which lets users register, create communities and forums and interact among them. For the last edition of the Day LINHD included also the bilingual plugin WPML to make it available the possibility of including translation in Spanish and English. This feature was, however, just used for the general website and its blog entries.

The data employed in this proposal has been obtained by using web scraping techniques (Fredheim, 2014) in the DayofDH websites for the previously mentioned editions. In particular, humanists' blog data, and their associated posts in the website have been gathered in this phase. All information scrapped from blogs is public and accessible from the Internet and, also, they have been anonymized for ethical issues. For validation purposes, the conceptual information about the database schemas have been compared with the extracted dataset, concluding the extraction process has been satisfactory.

Since the data obtained are huge enough to be efficiently processed, the use of big data techniques have been considered for this work (<http://social-metrics.org/analyzing-big-data-with-python-pandas/>). In order to achieve the main goals of the project, all the information related to the textual content in DayofDH have been processed, so that the most significant tokens are selected. Then, these tokens have been characterized by two parameters; first, it has been used the direct frequency which characterize if a token is used regularly in all DayofDH blogs. Secondly, the inverse frequency of the token that give information of how significant the token is in the context of digital humanities in a semantic way.

These parameters have been used to observe the interest and evolution of the characterized tokens along the time, either in a global and individual way. The interest of the global analysis is to find how the knowledge has evolved during the years of the study. From the point of view of a personal analysis, the interest is to build individual profiles that show the main interest of the researchers in the humanist community. Finally, the leadership relations have also been explored by using disease propagation techniques in the generated social network, taken into account the different editions of the DayofDH. For instance, Fig. 1 shows the social network according to the amount of authors' participations.

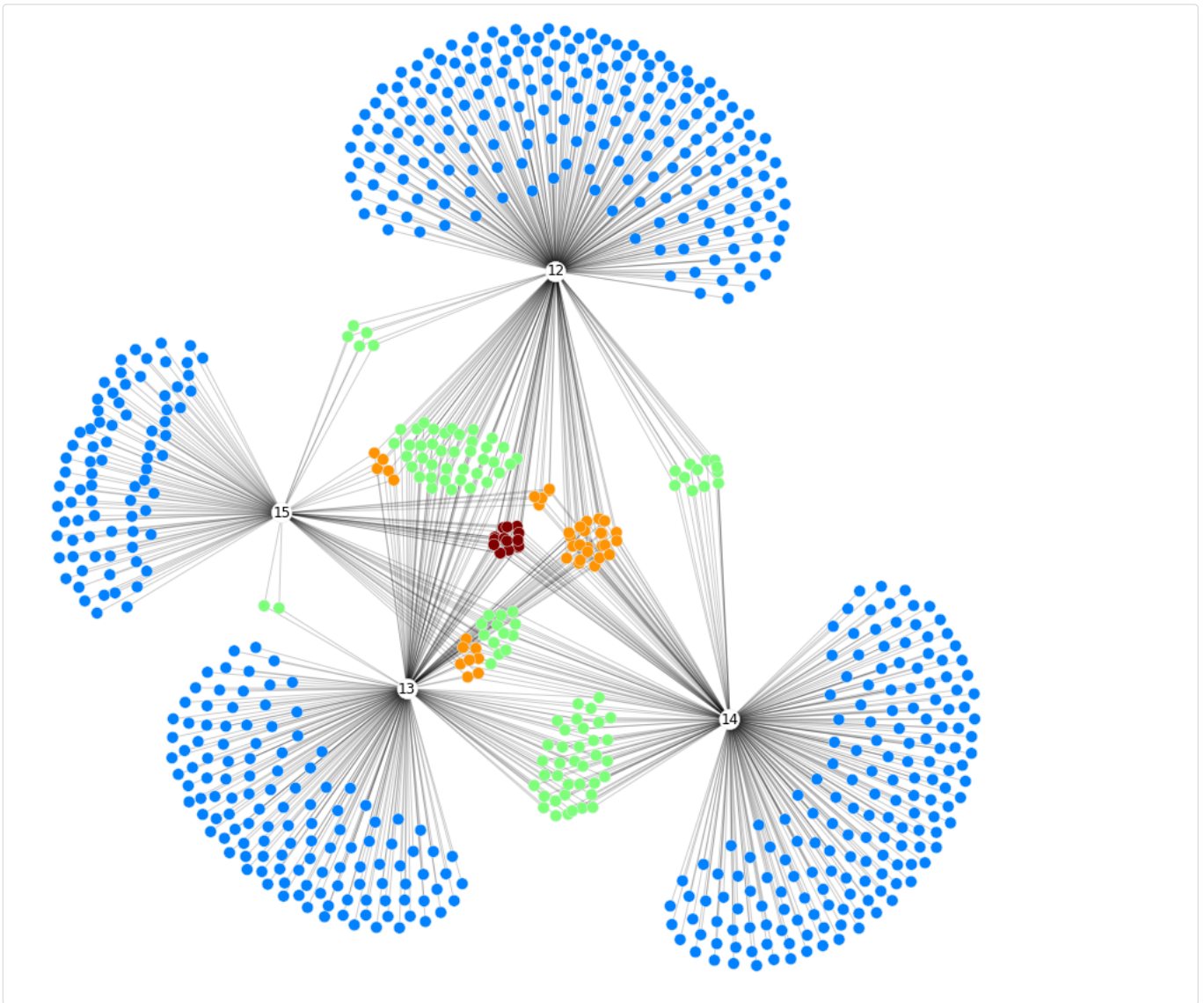


Fig. 1. Social network generated for 2012-2015 editions of DayofDH

The resulting graphics and visualizations (Barra et al., 2014b) let users make a quick idea of how the DH focus has been moving and distributing across the time through the different Academies in the different countries, but also how topics and interests change from one country to another and it is strongly related to their perspectives and disciplines, which are not independent from their origin (as an example, see Figs. 2 and 3). This approach will enlighten future studies on DH perspectives with real and precise data on the current state-of-the-art on DH perception and its evolution. Data of the years 2009, 2010 and 2011 are not used at the moment, as the same information is not available through web scraping. They will be incorporated to this study as a future work.



