

Saluteca, a Spanish Health Video Portal

Sánchez-Bocanegra C.L.^{#1}, Rivero A.^{*2}, Fernández-Luque L.^{#3}, Sevillano J.L.^{*4}

[#]*Department of Computer Science, University of Seville, Spain*

⁴jlsevillano@us.es

^{*}*Northern Research Institute- Tromsø, Norway*

¹carlosl.sanchez@atc.us.es

²alejandro.rivero@norut.no

³luis.luque@norut.no

Abstract— Videos are the most common, direct and easy way to transmit an idea or a message and they are one of the more popular ways people use to create and upload contents on the Internet. Youtube is a resource that makes this possible, but one of its weak points is that it hosts many mixed videos which are of low interest and can be considered as “spam”.

On the other hand, the use of Social Media is increasing in the health sector, as patients and health professionals need to answer questions and express their experiences and feelings. The lack of health information encouraged these actors to turn to social media for support and advocacy, and videos have the advantages of being engaging and easy to understand.

Saluteca is a web platform that collects trustworthy Spanish health videos from Youtube and offers them in a meaningful way. This paper describes Saluteca and explains the crawling method used to retrieve these videos from trusted Spanish channels.

I. INTRODUCTION

According to Kietzmann et.al, Social Media are "interactive platforms via which individuals and communities create and share user-generated content" [1]. Nowadays, Internet boosts Social Media with a continued growth [2, 3, 4]. People want to share their feelings, experiences and knowledge in all kind of formats: blogs, podcasts, videos... [3, 4].

Thus videos are one of the possibilities of Social Media. They are easy to record [5] and publish [6] on any device such as a smartphone which use is more extended every day.

Over 76% percent of surveyed U.S. hospitals and health systems said they use the social networking website Facebook, [according to a report](#) by consultancy firm CSC. The report also states that 65% use Youtube [7].

Youtube is a video-share website [8] that gives us the opportunity to upload and view them. Since 2005, Youtube is considered the largest and most popular video community which includes user-generated contents [9].

Users can watch and share their videos in this service web. We use this cloud service and get these resources which are loosing in an transparent abstract layer [10].

Youtube contains: videos, which are the main content of this web; channels, which are useful for organizations and companies allowing gathering a group of videos; playlists, to gather videos in a list with similar characteristics; and users, who can include their own videos, and they can also personalize and configure all their video metadatas. Tags, categories, description, geolocation, etc, are some examples of metadata.

Near 800 million unique users access Youtube, 72 hours of video are uploaded every minute, over 4 billion hours of video are watched each month, 100 million take a social action on this platform (likes, share, comments) [11].

Youtube includes a very useful documentation with many examples that help developers to create new code [12]. This makes quite differences in the quality of final results.

Besides, Youtube offers a comprehensive and complete API to personalize webs [12]. This API is available to retrieve all metadata and to embed video properties inside webs. Users can configure these properties to let developers to export their videos. It defines the way to access data and customize videos. Developers use Youtube API to integrate it into their website, to create mashups [13, 14] and even to integrate it in mobile systems [15].

The main objective of this study is to select interesting videos according to HealthTrust research [16], collecting all videos from Spanish Youtube Channels and Playlists in a Drupal platform and provide a web-based system to show these metadata and videos.

Drupal is a robust and powerful content management system (CMS) increasingly used in health communities [17, 18] that overwhelms functions and procedures including a personalized image and video. It is an open source application with huge communities contributing to their improvements.

Given the huge amount of videos available in Youtube, users may feel overwhelmed with too many contents that are usually full of “spam” (advertising, misleading videos, etc.) and other kind of useless videos. In addition users can spend too much time trying to get good videos relevant for their diseases [19].

Saluteca gathers a set of well-known channels and retrieves their videos, thus users can select a great number of videos from these trusted channels. Saluteca provides a very simple solution to a simple problem: Users need to find trusted videos from Internet. This module is a complement to the Video module that allows you to store videos locally on your site and to refer to remote videos from Youtube and Vimeo.

II. METHOD

We need a way to store videos using Web 2.0 platforms. In this way Saluteca features include an open source and social engineering release developed by the Northern Research Institute (Norut) and users will jointly identify videos in a controlled and secure platform.

Saluteca uses Youtube API together with Drupal CMS and crawler looks into the 22 trusted channels: Fundación del corazón, Asociación Española Contra el Cáncer, Esclerosis múltiple, Mundo bebe tv, Feder, El blog de Rosa, Ministerio

de Salud del gobierno de Chile, Proyecto salvavidas, Ministerio de la Salud de la Nación, Torre vieja Salud, Fundación Josep Carreras contra leucemia, Escuela de Pacientes, Noscuidamos, Consejería de Salud de la Junta de Andalucía, Paciente Experto, Fundiabetes, Hospital San Joan de Déu, Fundación de ayuda contra la drogadicción, Madrid Salud, Salud Andalucía, Nestle Tv Bebe, Sanidad Cantabria, Ministerio de Sanidad, Servicios Sociales e Igualdad.

Before we start, we need a PHP server configuration [20, 21] (Apache, PHP setup files), Youtube API client Library [22], Drupal content type definitions that wrap youtube video entity, state type definitions to store every transition in drupal and time-based schedule (cron) to relaunch execution on each iteration.

Crawler automatically scans Youtube resources to select videos and insert them into the Saluteca web. It explores every video extracting titles, descriptions, keywords, references, categories, and other metadata, it reports back to Drupal API framework nodes and adds the information to their huge database. Finally, These nodes include them into the backbone using search engines combined with clever algorithms.

Therefore, the basic crawler steps are: retrieving the last state from the last indexing, processing 200 items video nodes on each iteration (100 to update videos from our channel with Youtube and 100 to insert new Youtube videos) and finally saving the new states to the system keeping all ready for next iteration (see Figure 1).

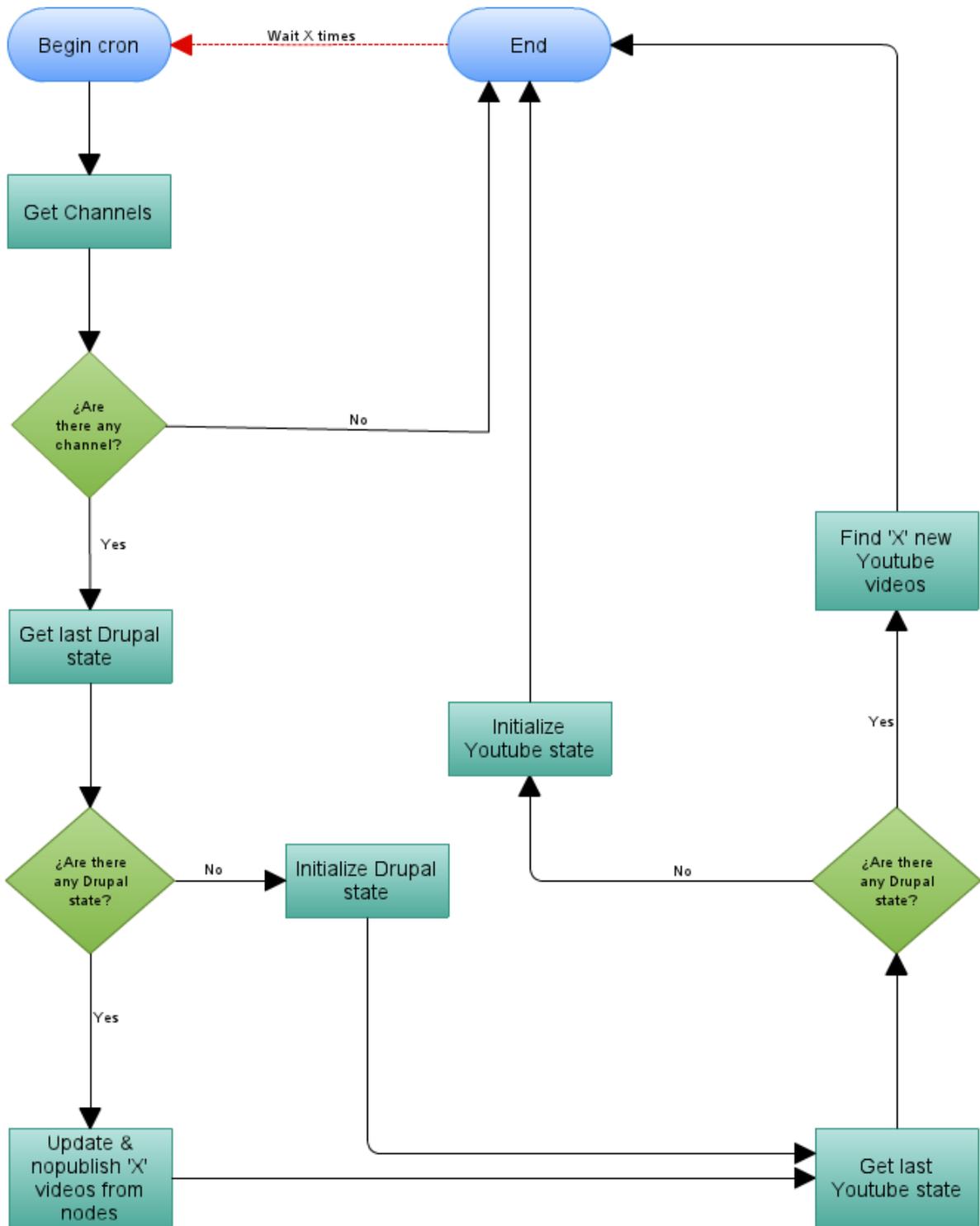


Figure 1: Main Crawler flowchart

Drupal and Youtube setups let us to show and customize videos in different ways: detail pages with simply thumbnail listings, full embed video, etc. It also provides support for retrieving and displaying thumbnail images for each video.

The system maintains a mapping between the contents, the metadata associated with the individual pieces of content, and the URL of the source from which the content was crawled.

III. RESULTS

The results include metadata and video properties (tags, categories, description) together with the embedded video.

Our collected samples are just a fraction of the data provided by YouTube database. Saluteca offers one solution to share experiences and knowledge in a learning process between users.

We get over 1437 videos from 22 spanish channels and the crawler updates and inserts 200 videos on each iteration (1 hour). Besides, the system automatically deletes the videos selected by the authors. Figure 2 and 3 show some video example previews.

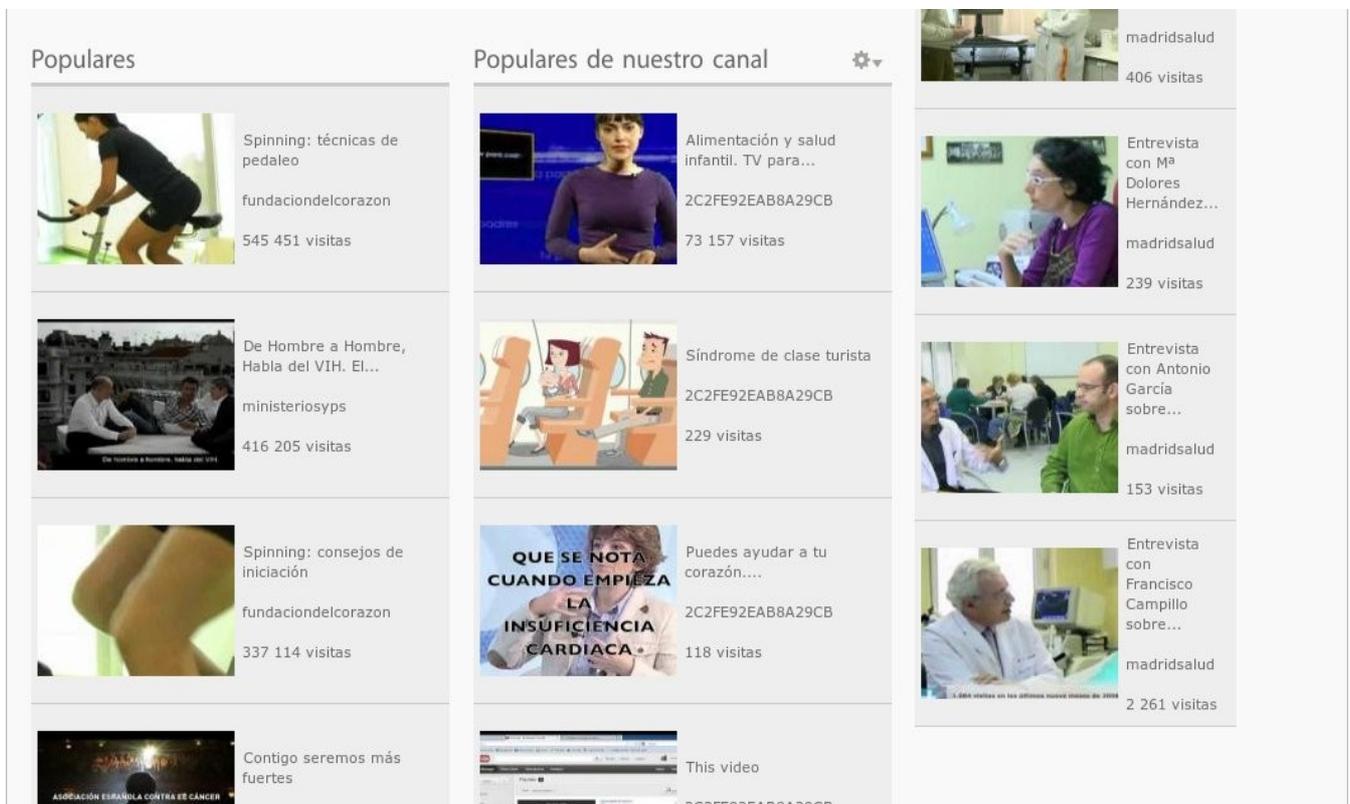


Figure 2: Saluteca preview image

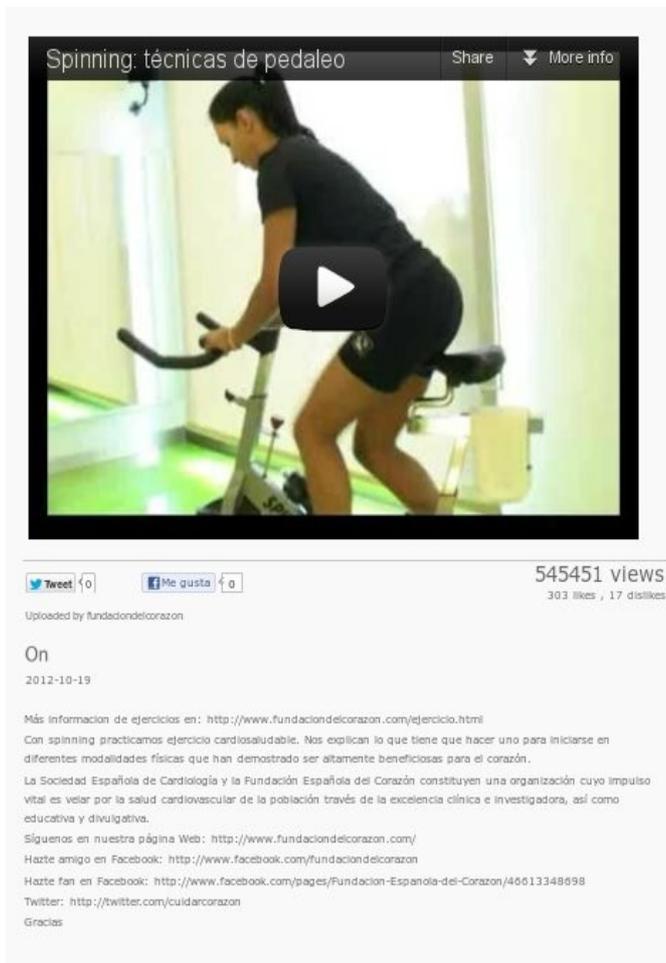


Figure 3: Saluteca preview image

IV. DISCUSSION

Youtube provides an easy way to produce mashups and use the potential and use the potential involved in audiovisual communication. We can find many Google groups who give support and quick answers too.

Youtube API [11] is well documented with good examples, however the API threshold is rather ambiguous. A limitation is the service quota: it is not properly documented and its criteria are being modified continuously. To verify the right performance of the API service, we tested the iteration time in different ranges, as well as different amounts of videos to process.

Drupal is a stable and well documented content management system. It has forums with thousands of helpful topics [23] and over 10,000 developers show are contributing with the source code [24]. Drupal has multiple options to configure it manages the video platform. It supports embedded videos from any provider (such as Youtube or Vimeo).

Experiments on a real life dataset from our videos would demonstrate that recommender system overwhelms a better video suggestions [25], it will also be able to recommend related videos.

ACKNOWLEDGMENT

REFERENCES

- [1] Kietzmann, J.H., Hermkens, K., McCarthy, I.P., & Silvestre, B.S. (2011). "Social media? Get serious! Understanding the functional building blocks of social media" (PDF). *Business Horizons* 54 (3): 241–251. doi:10.1016/j.bushor.2011.01.005. Retrieved September 17, 2012
- [2] Van de Belt TH, Berben SA, Samsom M, Engelen LJ, Schoonhoven L. Use of Social Media by Western European Hospitals: Longitudinal Study. *Journal of Medical Internet Research*. 2012 May 1;14(3):e61.
- [3] Social Media and Young Adults from Pew Internet <http://www.pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/Part-2/1-Cell-phones.aspx> available at <http://www.webcitation.org/6Bj9AN7mu>
- [4] Teens and Social Media from Pew Internet <http://www.pewinternet.org/Reports/2007/Teens-and-Social-Media/1-Summary-of-Findings/1-Summary-of-Findings.aspx> available at <http://www.webcitation.org/6Bj9Mrpk1>
- [5] Gold J, Pedrana AE, Stooze MA, Chang S, Howard S, Asselin J, et al. Developing Health Promotion Interventions on Social Networking Sites: Recommendations from The FaceSpace Project. *Journal of Medical Internet Research*. 2012 Feb 28;14(1):e30.
- [6] Fernandez-Luque L, Karlsen R, Melton GB. HealthTrust: A Social Network Approach for Retrieving Online Health Videos. *Journal of Medical Internet Research*. 2012 Jan 31;14(1):e22.
- [7] Which Types of Social Media Platforms Do Surveyed Hospitals and Health Systems Use? <http://www.ihealthbeat.org/data-points/2012/which-types-of-social-media-platforms-do-surveyed-hospitals-and-health-systems-use.aspx>
- [8] Chou W-YS, Hunt Y, Folkers A, Augustson E. Cancer Survivorship in the Age of YouTube and Social Media: A Narrative Analysis. *Journal of Medical Internet Research*. 2011 Jan 17;13(1):e7.
- [9] Cheng X, Dale C, Liu J. Statistics and Social Network of YouTube Videos. 16th International Workshop on Quality of Service, 2008. IWQoS 2008. 2008. p. 229–238.
- [10] Burgess JE, Green JB. YouTube: online video and participatory culture [Internet]. 2009 [cited 2012 Oct 27]. Available from: <http://www.polity.co.uk/book.asp?ref=9780745644783>
- [11] Youtube statistics http://www.youtube.com/t/press_statistics available at <http://www.webcitation.org/6BjAbpLHR>
- [12] Youtube API <https://developers.google.com/youtube/> available at <http://www.webcitation.org/6BjA1bzCn>
- [13] Benslimane D, Dustdar S, Sheth A. Services Mashups: The New Generation of Web Applications. *IEEE Internet Computing*. 2008 Oct;12(5):13–15.
- [14] Wong J, Hong JI. Making mashups with marmite: towards end-user programming for the web. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* [Internet]. New York, NY, USA: ACM; 2007 [cited 2012 Oct 27]. p. 1435–44. Available from: <http://doi.acm.org/10.1145/1240624.1240842>
- [15] Fox R, Cooley J, McGrath M, Hauswirth M. Mobile health apps - from singular to collaborative. *Stud Health Technol Inform*. 2012;177:158–63.
- [16] Fernandez-Luque L, Karlsen R, Melton GB. HealthTrust: trust-based retrieval of you tube's diabetes channels. *Proceedings of the 20th ACM international conference on Information and knowledge management* [Internet]. New York, NY, USA: ACM; 2011 [cited 2012 Oct 27]. p. 1917–20. Available from: <http://doi.acm.org/10.1145/2063576.2063854>

- [17] Papanicolaou A, Heckel DG. The GMOD Drupal bioinformatic server framework. *Bioinformatics*. 2010 Dec 15;26(24):3119–24.
- [18] Ficklin SP, Sanderson L-A, Cheng C-H, Staton ME, Lee T, Cho I-H, et al. Tripal: a construction toolkit for online genome databases. *Database (Oxford)*. 2011;2011:bar044.
- [19] Fernandez-Luque L, Karlsen R, Bonander J. Review of Extracting Information From the Social Web for Health Personalization. *Journal of Medical Internet Research*. 2011 Jan 28;13(1):e15.
- [20] PHP language <http://www.php.net/> available at <http://www.webcitation.org/6BjBuOtR3>
- [21] Drupal CMS <http://drupal.org> available at <http://www.webcitation.org/6BjBxJCeW>
- [22] Youtube API client library https://developers.google.com/youtube/2.0/developers_guide_php
- [23] Drupal forum topics <http://drupal.org/forum> available at <http://www.webcitation.org/6BvmBVnQ9>
- [24] Statistics values and comparative of differentes content management systems <http://visual.ly/junk-trunk> available at <http://www.webcitation.org/6Bvml6dBW>
- [25] Jäschke R, Marinho L, Hotho A, Schmidt-Thieme L, Stumme G. Tag Recommendations in Folksonomies. In: Kok J, Koronacki J, Lopez de Mantaras R, Matwin S, Mladenic D, Skowron A, editors. *Knowledge Discovery in Databases: PKDD 2007* [Internet]. Springer Berlin / Heidelberg; 2007 [cited 2012 Oct 27]. p. 506–14. Available from: <http://www.springerlink.com/content/h734wqn430821546/abstract/>