

# Sistema de analítica de información con SPARQL para grafos de DBPEDIA.

Oswaldo Moscoso-Zea<sup>1</sup>, Edison Morocho<sup>1</sup>  
Institución: Universidad UTE<sup>1</sup>

Autor para correspondencia: Oswaldo Moscoso<sup>1</sup>, omoscoso@ute.edu.ec



# DBpedia

Figura 1. Logo DBpedia

## INTRODUCCIÓN

Con los años, Internet se ha convertido en un medio para almacenar, transferir y difundir información. En esta era del conocimiento, Internet también ha evolucionado hasta convertirse en una plataforma interconectada que contiene una gran cantidad de información de diferentes campos. Estas interconexiones pueden ser fuente de conocimientos muy interesantes que aún no se explotan. Por otro lado, uno de los problemas de esta vertiginosa evolución de millones de dispositivos conectados que hoy se conocen como Internet de las Cosas es la sobrecarga de información que tienen las organizaciones y gobiernos, lo que dificulta su correcto procesamiento y gestión. Uno de los campos que ha evolucionado para interconectar datos estructurados de diferentes fuentes es Linked Data. Linked Data es el nombre que se le ha dado al método que permite conectar datos de diferentes sitios web, para generar nuevo conocimiento a partir de estas conexiones. Uno de los proyectos emblemáticos en esta área es DBpedia, que es un conjunto de datos obtenidos de Wikipedia para proponer una Web Semántica que permite el análisis automático de las relaciones existentes entre los datos estructurados creados en Wikipedia, está interconectado con otros sitios como Geonames.

En DBpedia, el acceso a la información no puede ser consultado por usuarios normales y el modo de consulta es a través del "editor de consultas SPARQL virtuoso". Esta es una plataforma completamente gratuita y accesible para las personas que saben cómo manejar consultas SPARQL. La información que se muestra en el "virtuoso editor de consultas SPARQL" se muestra a través de enlaces de acuerdo con los criterios de búsqueda propuestos. Esta información no es muy detallada a la hora de realizar consultas ya que las búsquedas se realizan mediante parámetros específicos.

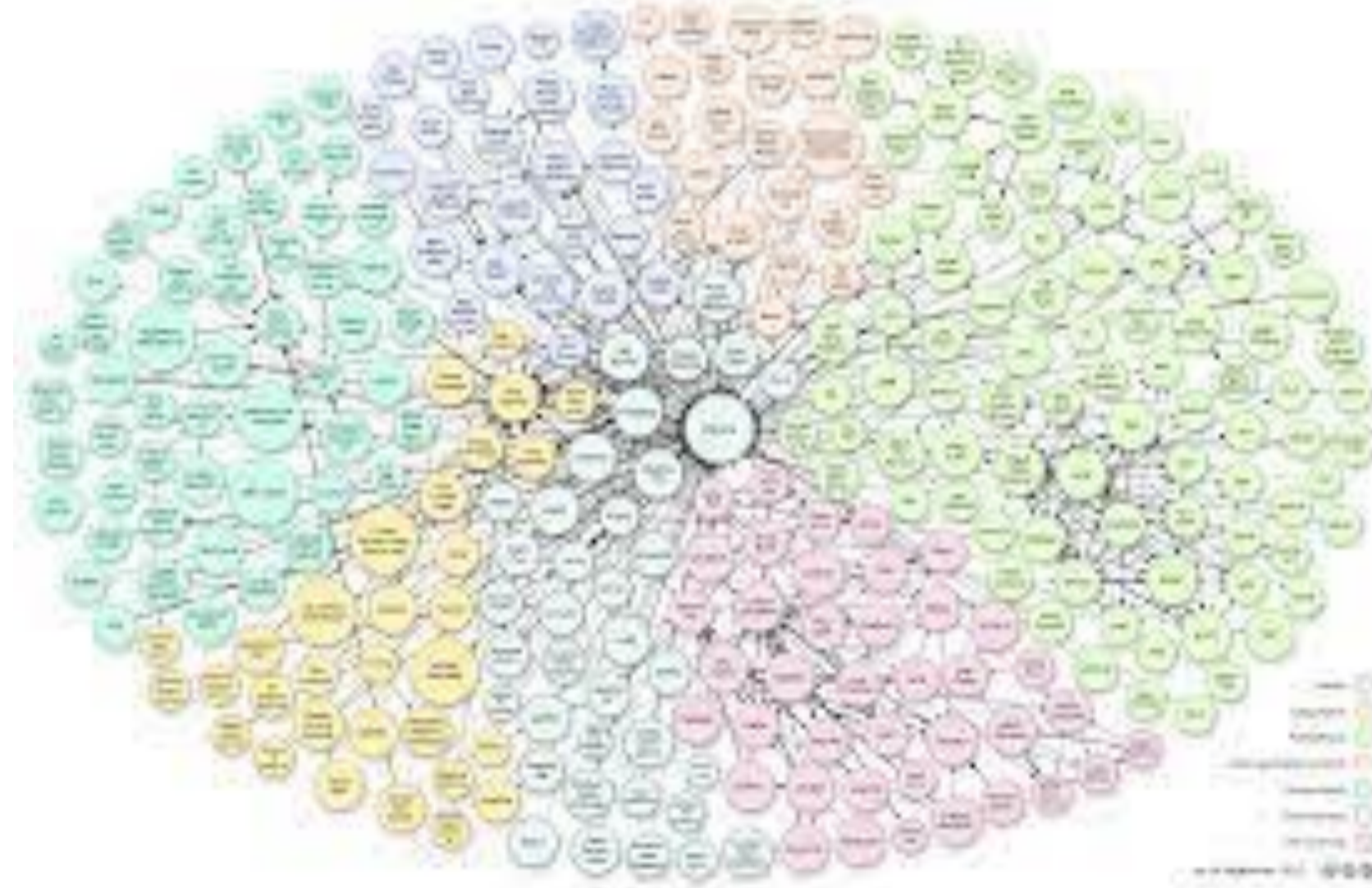


Figura 2. Datos enlazados en DBpedia

## MATERIALES Y MÉTODOS

El sistema informático desarrollado permite extraer información de la base de datos DBpedia usando el lenguaje de consulta SPARQL, la metodología usada es SCRUM. Esta metodología es implementada para tener un mejor manejo de las actividades y de los procesos de construcción de dicho sistema. La presente investigación es de tipo exploratorio-descriptivo. La investigación exploratoria sirvió para aclarar la naturaleza exacta del problema y para tener un primer acercamiento al objeto de estudio con una temática nueva en el país. El aplicar este método ayudó a tener una visión y conocimiento general del tema a tratar, de esta manera se realizó los estudios iniciales y establecer los alcances finales del proyecto.

## CONCLUSIONES

- La implementación del sistema permite a usuarios e investigadores contar con una primera visión del sistema (prototipo) que permite la extracción de información desde la DBpedia mediante el lenguaje SPARQL.
- El diseño y desarrollo del sistema BUSCAPEDIA, busca ser una herramienta para usuarios que no posean conocimientos del lenguaje de consulta SPARQL, y que puedan interactuar con la información que maneja DBpedia.
- El uso de la metodología SCRUM permitió el desarrollo rápido del sistema. De la misma forma, se pudo aplicar todas las ventajas que dicha metodología otorga al momento de su selección.
- El lenguaje de consulta SPARQL ayudó con la implementación de linked data, haciendo que el sistema extraiga la información de la manera más dinámica posible y manteniendo dos parámetros de búsqueda.
- La implementación del sistema BUSCAPEDIA (prototipo) es una herramienta con posibilidades de mejora, en la parte de búsqueda y recolección de datos. Si bien es cierto la herramienta BUSCAPEDIA no compensa el 100% de búsquedas realizadas por los usuarios. Se podría desarrollar o implementar un algoritmo, que permita al sistema reconocer las entidades que posee la base de datos DBpedia, con la finalidad de conseguir mejores resultados

## RESULTADOS Y DISCUSIÓN

El presente trabajo se centró en la realización de una investigación aplicada que comprenda el análisis, diseño e implementación de un prototipo de motor de búsqueda que extrae información parametrizada de la base de datos DBpedia.

La novedad de este proyecto de investigación es que el sistema propuesto ha sido desarrollado para permitir la interacción dinámica a través de una interfaz con los datos DBpedia almacenados.

La implementación del sistema se basó en realizar un software que ayude a extraer la información de la base de datos DBpedia, haciendo uso del lenguaje de consulta SPARQL. El software diseñado está pensado en los usuarios que no están familiarizados con el uso y manipulación de SPARQL y los datos de DBpedia. El sistema cuenta con un pequeño manual para los usuarios que especifica cuáles son los parámetros de búsqueda. Además, el sistema cuenta con tres lenguajes (español, inglés, francés), cada uno de ellos contiene diferente información a ser procesada, esto se debe a que no toda información se encuentra en un solo lenguaje.

El sistema controla dos parámetros de búsqueda. Esto hace que el sistema sea más dinámico a la hora de manipular y visualizar los datos ingresados por el usuario. El sistema muestra la consulta SPARQL como una ventana emergente



Figura 3. Sistema de búsqueda original Virtuoso Sparql

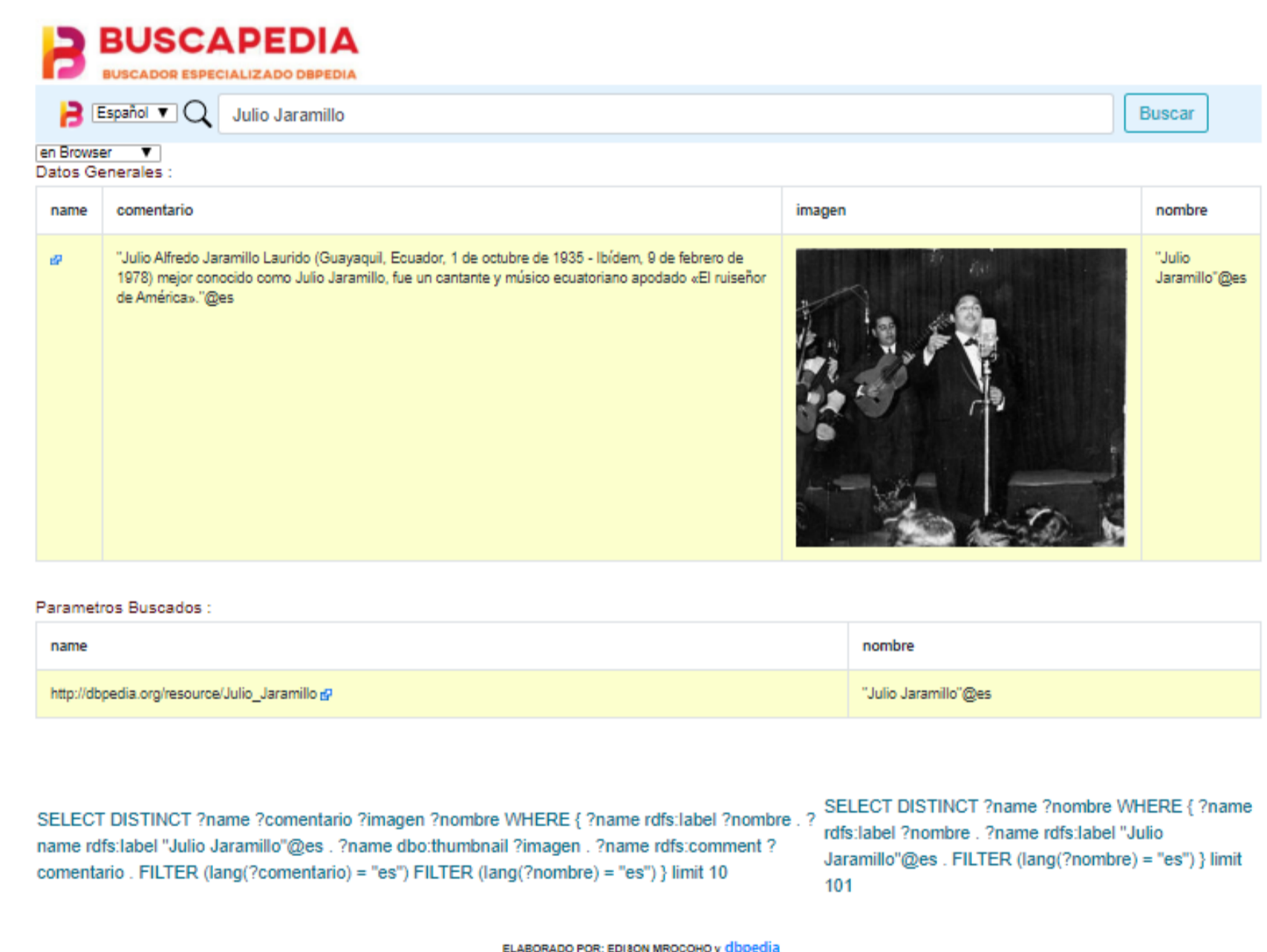


Figura 4. Sistema de búsqueda desarrollado en DBpedia - Buscapedia

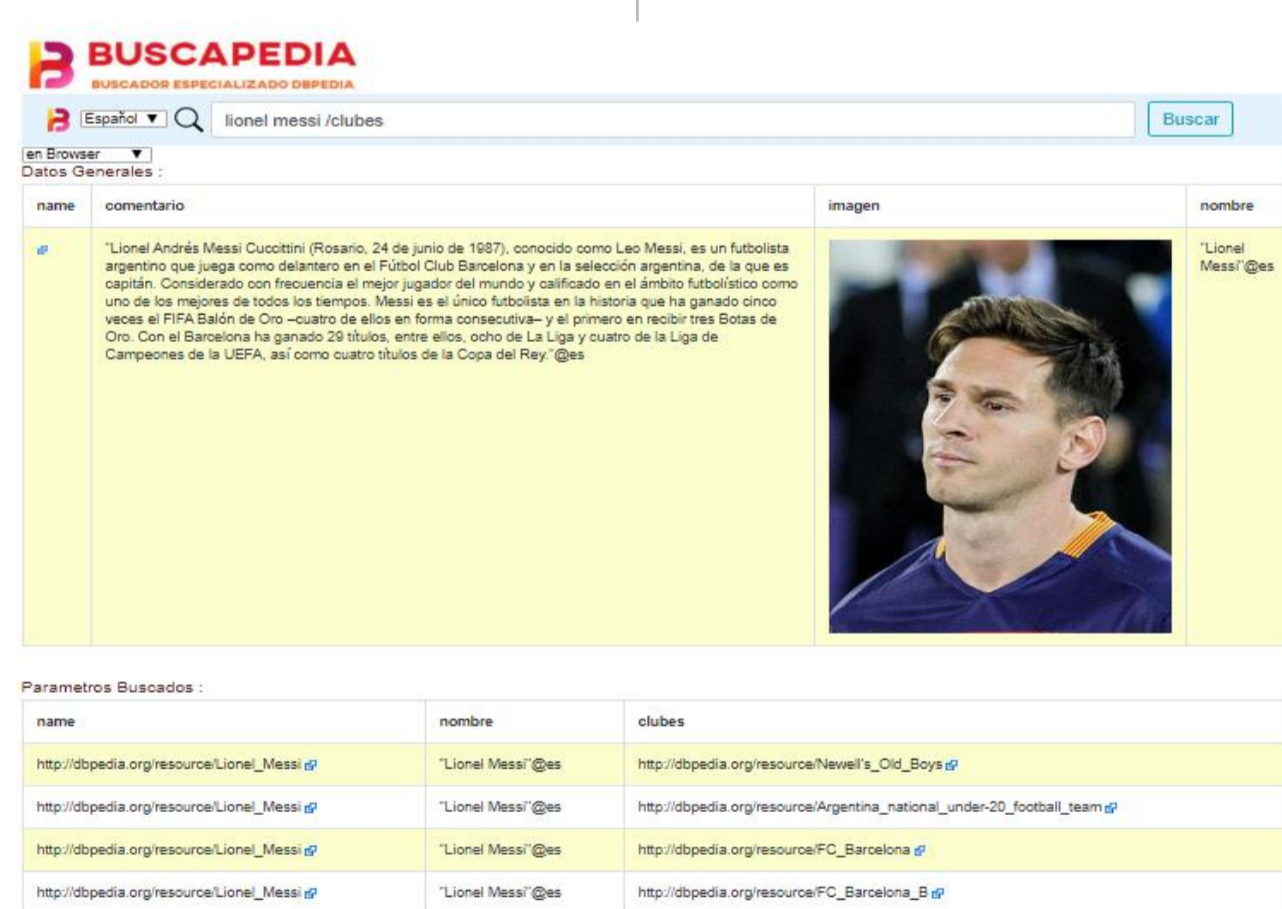


Figura 5. Sistema desarrollado ejemplo de búsqueda

## REFERENCIAS

Adolfo, T. L. (2001). Ontologías en la web semántica. *Departamento de Informática de La Universidad De*, 1-4.

Beniwal, R., Gupta, V., Rawat, M., & Aggarwal, R. (2018). Data Mining with Linked Data: Past, Present, and Future. *Proceedings of the 2nd International Conference on Computing Methodologies and Communication, ICCMC 2018*, (Iccmc), 1-5. <https://doi.org/10.1109/ICCMC.2018.8487861>

Bizer, C., Cyganiak, R., & Gauß, T. (2007). The RDF book mashup: From Web APIs to a Web of data. *CEUR Workshop Proceedings*, 248.

Britto Montoya, J. A. (2016). Comparación de metodologías ágiles y procesos de desarrollo de software mediante un instrumento basado en CMMI Mapping agile methodologies and software development processes using a CMMI based. *Scientia et Technica*, 21(2), 150-155.

Das, N. N., Chowdhary, M., Luthra, R., Maisera, M., & Garg, S. (2019). Semantic Big Data Searching in Cloud Storage. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*, 351-355. <https://doi.org/10.1109/COMITCon.2019.8862188>

DBpedia. (2016). Acerca de DBpedia.

Feridun, M., & Tanner, A. (2010). Using linked data for systems management. *Proceedings of the 2010 IEEE/IFIP Network Operations and Management Symposium, NOMS 2010*, 926-929. <https://doi.org/10.1109/NOMS.2010.5488341>

Jontokostas, D., Bratsas, C., Auer, S., Hellmann, S., Antoniou, I., & Metakides, G. (2012). Internationalization of Linked Data: The case of the Greek DBpedia edition. *Journal of Web Semantics*, 15, 51-61. <https://doi.org/10.1016/j.websem.2012.01.001>

Kehmann, J., Isele, R., Jakob, M., Jentsch, A., Kontokostas, D., Mendes, P. N., ... Bizer, C. (2015). DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2), 167-195. <https://doi.org/10.3233/SW-140134>

Phyue, S. L., Thein, M. M., Win, T. T., & Thwin, M. M. S. (2010). Semantic web information retrieval in XML by mapping to RDF schema. *ICNIT 2010 - 2010 International Conference on Networking and Information Technology*, 500-503. <https://doi.org/10.1109/ICNIT.2010.5508464>

Quilitz, B., & Leser, U. (2008). Querying distributed RDF data sources with SPARQL. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5021 LNCS, 524-538. [https://doi.org/10.1007/978-3-540-68234-9\\_39](https://doi.org/10.1007/978-3-540-68234-9_39)

Sowe, S. K., & Zettsu, K. (2016). Towards an Open Data Development Model for Linking Heterogeneous Data Sources. *Proceedings - 2015 IEEE International Conference on Knowledge and Systems Engineering, KSE 2015*, 1-4. <https://doi.org/10.1109/KSE.2015.56>

Sowmya, R., & Suneetha, K. R. (2017). Data Mining with Big Data. *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*, 26(1), 246-250. <https://doi.org/10.1109/ISCO.2017.7855990>

Strawn, G., Andreessen, M., & Bina, E. (2014). Masterminds of the web. *IT Professional*, 16(4), 58-59. [https://doi.org/10.1016/S0169-7552\(98\)00110-X.A](https://doi.org/10.1016/S0169-7552(98)00110-X.A)

Thuy, P. T. T., Lee, Y. K., Lee, S., & Jeong, B. S. (2008). Exploiting XML schema for interpreting XML