

Best Practice Presentation: **Ecco!: A Linked Open Data Service for Collaborative Named Entity Resolution**

Matthew Miller
New York Public Library,
NYPL Labs,
United States
matthewmiller@nypl.org

M. Cristina Pattuelli
Pratt Institute, School of
Information and Library
Science,
United States
mpattuel@pratt.edu

Keywords: named entity resolution; Linked Open Data (LOD); Ecco!;

Abstract

This demo proposal presents Ecco!, a Linked Open Data (LOD) application for entity resolution. Specifically, Ecco! is designed to disambiguate and reconcile named entities with URIs from authoritative sources. Technically, Ecco! creates a wrapper around LOD APIs of suitable datasets such as VIAF and Freebase to retrieve data useful for supporting entity matching. The system automatically ranks and groups the results into different clusters according to various confidence levels – from exact matches to one to many or no matches. The quality of the data output can be further refined through human disambiguation consisting of validating a match or identifying the correct URI when multiple matches are possible.

Ecco! is designed to enable users to quickly and easily contribute to this curation process. The system provides an intuitive user interface that supports a collaborative workflow where a community can work together in a distributed and incremental way. The combination of automated matching plus human curation has the potential to produce a superior quality of data, not currently achievable through traditional methods.

This application works alongside existing legacy systems and data sources through an import and export workflow. Extracts generated from a legacy system or data source are enriched through Ecco! and then looped back to update the originating source. Ecco! intends to address the well-known "bucket names" problem that occurs when legacy data has accumulated and contains a mix of heterogeneous names derived from different authorities (e.g., LC/NAF, ULAN, etc.) as well as locally defined terms.

Ecco! is a node.js application that anyone can download and run on their local system. There is no need for a server installation, but it could be installed on a server to allow for the collaboration of an unlimited number of participants. Ecco! has the capacity to work with LOD APIs in a modular way. While the demo version will specifically leverage VIAF and Freebase, any API plugin could be virtually written for it. Also, while in the current release the application will be centered on persons and organizations, other types of entities including geographic locations, events, topics, etc. could be also handled by the system.

Even though Ecco! was developed as part of the Linked Jazz project,¹ it is domain-agnostic and thus not tied to any specific context of use. The demonstration includes different scenarios showing a series of use cases. Results from a first round of testing will also be shared.

Data quality poses a daunting challenge in Linked Open Data development and requires the creation and adoption of new methods and tools to promote accuracy and consistency of data. Ecco! includes a series of innovative features that make it uniquely flexibility and easy to use.

¹ <http://linkedjazz.org>

Most notably, this system lowers the barrier for non-programmers who want to actively contribute to the production of high quality linked data through a user-friendly and collaborative platform.