Artangled – A Social Network of Artists

Dinesh Monteiro, Deepak Karkala, Monal Vora

University of Southern California
Computer Science Department
Los Angeles, CA 90089
{dmonteir, karkala, mvora}@usc.edu

1   Abstract

Artists all over the world might have had influences over each other and they might share some common relations. For example an Artist may be an assistant to another artist or may be a teacher of another artist. There may also be family relations like one artist being a grandfather, son, wife of another. Apart from these some other common relations include having same nationality, part of a same movement, interested in similar medium of art or subject of art etc. Information about artists and artworks is scattered on the web and it is not easily accessible to the user in a visually interactive way.

Our solution, Artangled is a collection of artists and their artworks that depicts a social network among them. This lures the user in exploring interesting relationships between artists of their choice and to also see the artworks of that artist.

Artangled helps the users to conveniently get answers to complex queries regarding relationships between the artists. For eg: If the user wants the details of all the artists who studied under the same person under whom Leonardo Da Vinci studied, he can easily get this using Artangled.

2   Motivation

Details about the relationships of various artists are not accurately documented anywhere on the web. Hence, there is a need for integration of the data from various sources on the web and consolidation of their results into one website that portrays all of it in a manner that is acceptable to the user so that he is interested into exploring further.

3   Approach

Figure 1 shows the overall architecture of Artangled. It lists all the steps involved in the process of building Artangled. Let us go through each step in detail.

Firstly, we extracted the data from our first source – Artcyclopedia[1]. We got the details of artists like name, birth year, death year, nationality, movement, medium etc. We also got the data of various relationships between artists like studied under, assistant to, father of, grandfather of, brother of etc. We then modelled the data into xml and gave the corresponding tags. This data needed cleaning as it had unnecessary tags like bold tags and some space characters. Also, the sources had life span data as a whole along with the birth and death year as (1961 - ?).We needed this to be separated as born year and died year as it would be further used while doing record linkage between the data of sources. Then we performed the necessary cleaning of the data using Karma[5], on open source tool developed by ISI at USC.
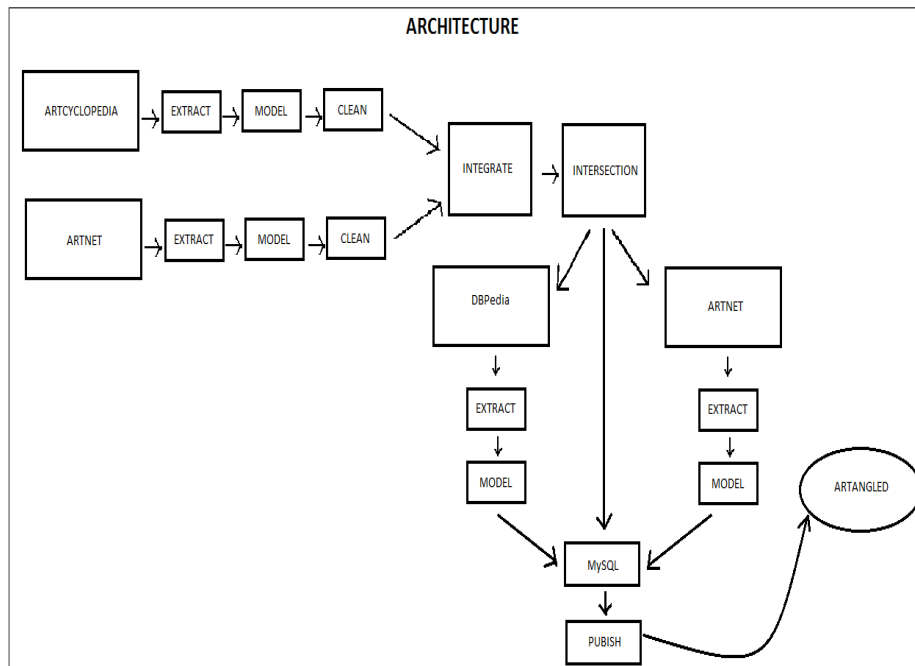

Fig. 1. Architecture Diagram

Secondly, we extracted data from our second source – Artnet[2]. Here we got the details of the artists like name, birth year, death year and nationality. In the modelling phase we put the data into an xml file. This data consisted of Unicode characters. Since this data was required for record linkage, we had to replace them with ascii equivalents of them. We accomplished this using Unidecode library available in Perl.

Next, we needed to do record linkage on the data obtained from the two sources. We used a tool called FRIL[4]. We did record linkage based on name, nationality, birth year and death year as these were the common details available in both the sources. The different weights given to the fields are listed in Table 1. If a field was missing 50% weight was still assigned to it. Also the Born and Died year were allowed for +/- 10 years mismatches in case if there would have been a typing error. The threshold for the records to match was kept 70%. We experimented with the dif-

ferent parameters and decided upon the best for our needs. We got a total of 6210 matching records from both the sources.

Field	Weight(%)
Name	40
Nationality	15
Born Year	25
Died Year	20

Table 1. Weight distribution during Record Linkage

We then got the images and a brief description about the artists that matched after record linkage from dbPedia[3] by using SPARQL library of .Net. We queried for dbpedia-owl:abstract for the description and foaf:depiction for the image of the artists. We could get the details of 3593 artists from dbPedia out of total 6210 records matched. This output was represented in the form of a CSV file.

Extracting artworks' details like title, image, year, medium, dimension was then done from Artnet. We extracted 579957 artworks. We limited each of the artists' artworks to 200. Then we exported the data to XML format.
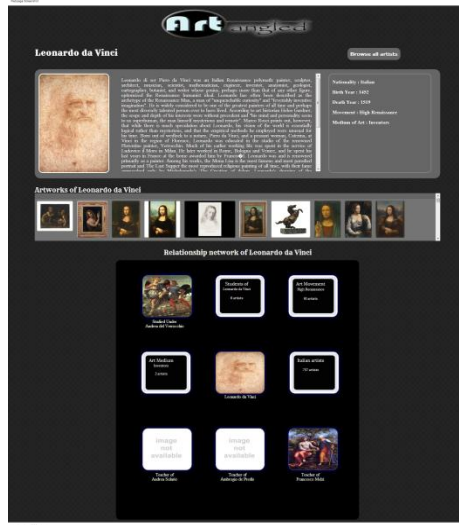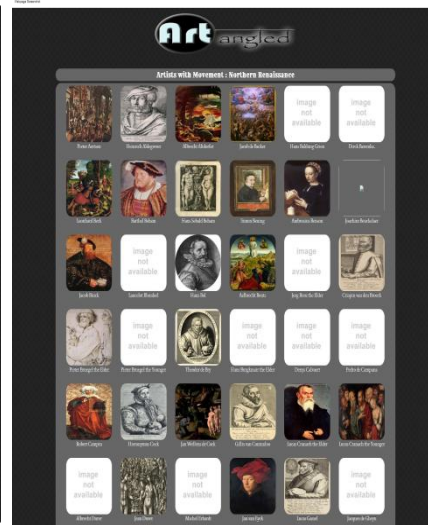

Fig. 2. Artist's Details Page

Fig. 3. Artists of the same category

Proceeding further we created a schema and populated MySQL database with the data from dbPedia, artworks details from Artnet and the artists' relationships and other details from the integrated data after record linkage.

Publishing the data onto the website involved using php as a backend to connect to the MySQL database and HTML, CSS, Javascript as frontend technologies to create the UI of the website "Artangled". Figure 2 shows the screenshot of the page display-

ing an artist's details. This page displays all the details of the artist along with the artworks and a network displaying the various relationships the artist is involved in. When we click on one of the relationships of that artist, for example the "Movement" relationship of an artist, we get a list of all other artists who are part of the same movement. Figure 3 shows screenshot of the page listing all the artists belonging to Art Movement "Northern Renaissance".

4   Empirical Results

During extraction phase, we extracted 8951 records for artists from Artcyclopedia. Whereas, we extracted 286,135 artists' records from Artnet.

After record linkage, off the 6210 records that matched, we took various random samples of 50 records and we could find that on an average around 47 records were relevant whereas there were about 2 to 3 false positives in the records matched.

When we extrapolate the results to total of 6210 records matched we could conclude that approximately 5900 relevant record matches were found. This gives a Precision of 0.95 which is pretty good. We have not calculated Recall for our results as finding missing records out of the 286,135 artists was not practical.

Figure 4 shows few examples of good matches that were found in the records linkage phase. We could see that cases like initials, slight mistakes in spelling, missing middle name were handled well by the algorithm. Even slight mismatch in birth and death years were accommodated.

Figure 5 shows some samples of false positives that were observed after record linkage. We could see that most of the false positives occurred because of lack of data for those records.

Source	Name	Nationality	Born	Died
Artcyclopedia	Aldro T. Hibbard	American	1886	1972
Artnet	Aldro Thompson Hibbard	American	1886	1972
Artcyclopedia	Antonietta Raphael Mafai	Italian	1895	1975
Artnet	Antonietta Raphael	Lithuanian	1895	1975
Artcyclopedia	Bernaert van Orley	Flemish	1488	1541
Artnet	Bernaert (Barend) van Orley	Flemish	1488	1541
Artcyclopedia	Ambrosius Holbein	German	1494	1519
Artnet	Ambrosius Holbein	German	1493	1526

Fig. 4. Record Linkage good matches

Source	Name	Nationality	Born	Died
Artcyclopedia	Alex Grey			
Artnet	Alex Green			
Artcyclopedia	Charles A. Meurer	American	1865	1955
Artnet	Charles A. Sommer	American		

Fig. 5. Record Linkage false positives

5 Conclusion and Future Work

"Artangled" successfully portrayed the relationships between artists with the help of a simple and clear social graph. It could successfully put together the various influencing relationships between artists that we initially had set out to document. We believe that the webpage would be a good stepping stone for further research in similar lines.

The way to go forward from here is that we can accommodate more artists in the network as well as explore more dimensions for the relationships. There is a scope for including various artworks and their relationships as well based on some of the common relations between them like Medium of art, Era, Region of origin etc.

6 References

1. http://www.artcyclopedia.com
2. http://www.artnet.com
3. http://www.dbpedia.org
4. http://fril.sourceforge.net/
5. http://www.isi.edu/integration/karma/