Language Technologies beyond research: from poetry to music industry

Elena González-Blanco
Global Head of Digital for Wealth Management and Insurance at Banco Santander, and Research Director at IE University
CARMINA BURANA

1. O Fortuna, velut Luna
   (O Fortune, variable as the moon)

2. Fortune plango vulnera
   (I lament Fortune's blows)

© Copyright 1937, 1963, 1967 by B. Schott's Sohn, Mainz
Printed in U.S.A.
The nostalgic approach to bring “millenials” onboard
HOW TO MAKE INSTAGRAM MUSIC VIDEO PREVIEWS

832 views • Liked by ycimaging and urban.scope

creativityanati: FINALLY After half of year waiting & waiting to release this amazing project, it finally dropped last night! @shooknightsffm ft. @mildurk2x #OBNOXIOUS
Covid impact: increase in time spent watching streaming videos

Source: Bloomberg (2020)
Music streaming

- Spotify: 32%
- Apple Music: 16%
- Amazon Music: 14%
- Tencent Music: 11%
- Google (inc YouTube Music): 6%
- Deezer: 2%
- Pandora: 1%
- Others: 1%

400 million music subscribers globally in Q1 2020

Source: MIDiA Research Subscriber Market Shares Model 06/20
17.000 M$ by 2024
Market annual growth rate of 5.8%

16% yearly growth of Latin Music
93% streaming services

½ of the 20 top songs in Spanish
10% of total market share
How do current systems recommend and classify songs?
Some AI experiments with voice and text
YELLOW SUBMARINE
Words and Music by Paul McCartney and John Lennon

In the town where I was born
Lived a man who sailed to sea
And he told us of his life
In the land of submarines.

So we sailed on to the sun
Till we found the sea green
And we lived beneath the waves
In our yellow submarine.

We all live in a yellow submarine
Yellow submarine, yellow submarine.
We all live in a yellow submarine
Yellow submarine, yellow submarine.

And our friends are all aboard
Many more of them live next door
And the band begins to play:
(Trumpets play or make the sounds yourself)

We all live in a yellow submarine
Yellow submarine, yellow submarine.
We all live in a yellow submarine
Yellow submarine, yellow submarine.

As we live a life of ease
Every one of us has all we need
Sky of blue and sea green
In our yellow submarine.

We all live in a yellow submarine
Yellow submarine, yellow submarine.
We all live in a yellow submarine
Yellow submarine, yellow submarine.
(Repeat chorus)
Diagram of the PoetryLab system:

- PoetryLab components:
  - Scansion
  - Hismetag
  - Enjambent

- Request for analyzing a Poem
- Request for querying Postdata Knowledge Graph
- Request for adding Metrical information
- Create your own Corpus
- Request for Building Postdata Knowledge Graph

- User Centric Front End

- AVERELL Corpus Creator
- Horace RDF creator
- StartDog Triple-Store

- Docker modules for each component.
Metrical analysis & scansion: Rantamplan

https://github.com/linhd-postdata/rantanplan
Ves ti das con man tos ne gro s
pien sa que el mun do es chi qui to
y el co ra zón es in men so .
Ves ti da con man tos ne gro s .
Pien sa que el sus pi ro tier no
y el gri to de sa pa re cen
en la co rrien te del vien to .
Ves ti da con man tos ne gro s .
Se de jó el bal cón a bier to
y el al ba por el bal cón
de sem bo có to do el cie lo .
¡Ay ya ya ya yay ,
que ves ti da con man tos ne gro s !
It includes over 86000 unique annotated poems from different languages.

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>lang</th>
<th>size</th>
<th>docs</th>
<th>words</th>
<th>granularity</th>
<th>license</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disco V2.1 (disco2_1)</td>
<td>es</td>
<td>22M</td>
<td>4088</td>
<td>381539</td>
<td>stanza line</td>
<td>CC-BY</td>
</tr>
<tr>
<td>2</td>
<td>Disco V3 (disco3)</td>
<td>es</td>
<td>2.8M</td>
<td>4080</td>
<td>377978</td>
<td>stanza line</td>
<td>CC-BY</td>
</tr>
<tr>
<td>3</td>
<td>Sonetos Siglo de Oro (adso)</td>
<td>es</td>
<td>6.8M</td>
<td>5078</td>
<td>466012</td>
<td>stanza line</td>
<td>CC-BY-NC 4.0</td>
</tr>
<tr>
<td>4</td>
<td>ADSO 100 poems corpus (adso100)</td>
<td>es</td>
<td>128K</td>
<td>100</td>
<td>9208</td>
<td>stanza line</td>
<td>CC-BY-NC 4.0</td>
</tr>
<tr>
<td>5</td>
<td>Poesía Lírica Castellana Siglo de Oro (plc)</td>
<td>es</td>
<td>3.8M</td>
<td>475</td>
<td>299402</td>
<td>stanza line word</td>
<td>CC-BY-NC 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>syllable</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Gongocorpus (gongo)</td>
<td>es</td>
<td>9.2M</td>
<td>481</td>
<td>99079</td>
<td>stanza line word</td>
<td>CC-BY-NC-ND 3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>syllable</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Eighteenth Century Poetry Archive (ecpa)</td>
<td>en</td>
<td>2400M</td>
<td>3084</td>
<td>2083688</td>
<td>stanza line word</td>
<td>CC-BY-SA 4.0</td>
</tr>
<tr>
<td>8</td>
<td>For Better For Verse (4b4v)</td>
<td>en</td>
<td>39.5M</td>
<td>103</td>
<td>41749</td>
<td>stanza line</td>
<td>Unknown</td>
</tr>
<tr>
<td>9</td>
<td>Métrique en Ligne (mel)</td>
<td>fr</td>
<td>183M</td>
<td>5081</td>
<td>1850222</td>
<td>stanza line</td>
<td>Unknown</td>
</tr>
<tr>
<td>10</td>
<td>Biblioteca Italiana (bibit)</td>
<td>it</td>
<td>242M</td>
<td>25341</td>
<td>7121246</td>
<td>stanza line word</td>
<td>Unknown</td>
</tr>
<tr>
<td>11</td>
<td>Corpus of Czech Verse (czverse)</td>
<td>cs</td>
<td>4100M</td>
<td>66428</td>
<td>12636867</td>
<td>stanza line word</td>
<td>CC-BY-SA</td>
</tr>
<tr>
<td>12</td>
<td>Stichotheque (stichopt)</td>
<td>pt</td>
<td>11.8M</td>
<td>1702</td>
<td>168411</td>
<td>stanza line</td>
<td>Unkwown</td>
</tr>
</tbody>
</table>
Adding a semantic layer

POSTDATA OntoPoetry V2

Controlled Vocabularies

POSTDATA Core

POSTDATA Poetic Analysis

POSTDATA Transmission

ODP

CIDOC CRM - FRBRoo
BERT vs ALBERTI multilingual model for poetry – similarity tools
Dalton: working on similarity

<table>
<thead>
<tr>
<th>Estopa - Como Camarón</th>
</tr>
</thead>
</table>
| Superior a mí, en la fuerza que me lleva, en el pulso que mantengo, con las oscuridad que tienen de oscuro tus ojos negros why que e cuentas el tiempo que pasa en tu pesañía why que me trae por esta calle de amargura why de lamento...

<table>
<thead>
<tr>
<th>Similarity based on: spacy information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cristal ahumado</td>
</tr>
<tr>
<td>Prima Tristeza</td>
</tr>
<tr>
<td>Abucheal</td>
</tr>
<tr>
<td>Brujeria</td>
</tr>
<tr>
<td>Mientras No Cueste Trabajo</td>
</tr>
<tr>
<td>Desidia</td>
</tr>
<tr>
<td>Insomnio de Amor</td>
</tr>
<tr>
<td>Poco a Poco</td>
</tr>
<tr>
<td>Tengo miedo</td>
</tr>
<tr>
<td>Similarity: 3.644</td>
</tr>
<tr>
<td>Similarity: 3.608</td>
</tr>
<tr>
<td>Similarity: 3.606</td>
</tr>
<tr>
<td>Similarity: 3.589</td>
</tr>
<tr>
<td>Similarity: 3.570</td>
</tr>
<tr>
<td>Similarity: 3.525</td>
</tr>
<tr>
<td>Similarity: 3.520</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Similarity based on: spacy_spotify information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volantes Pa La Falda Mi Gitana</td>
</tr>
<tr>
<td>Cristal ahumado</td>
</tr>
<tr>
<td>Brujeria</td>
</tr>
<tr>
<td>Nana de quebranto (Mala sombra)</td>
</tr>
<tr>
<td>Desidia</td>
</tr>
<tr>
<td>Insomnio de Amor</td>
</tr>
<tr>
<td>Poco a Poco</td>
</tr>
<tr>
<td>Similarity: 3.631</td>
</tr>
<tr>
<td>Similarity: 3.825</td>
</tr>
<tr>
<td>Similarity: 3.568</td>
</tr>
<tr>
<td>Similarity: 3.558</td>
</tr>
<tr>
<td>Similarity: 3.544</td>
</tr>
<tr>
<td>Similarity: 3.527</td>
</tr>
<tr>
<td>Similarity: 3.523</td>
</tr>
</tbody>
</table>
# POST /analysis

**Analyses a poem**

Gets enjambent and scanison info for the poem

## Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>operations</td>
<td>Choose which analysis is to be performed</td>
</tr>
<tr>
<td>array[string]</td>
<td>Default value: List [&quot;scansion&quot;, &quot;enjambment&quot;, &quot;entities&quot;]</td>
</tr>
</tbody>
</table>

- **scansion**
- **enjambment**
- **entities**

## Request body

Example Value: `text/plain`

```
Jamás encontraré más fiel ami-
go que en los peores momentos arrime
y no me trae mi verde
abriga que su cuerpo a mi alma abatida y anime
dándome el calor del mejor abrigo.
```
LYRAICS: ARTIFICIAL INTELLIGENCE FOR LYRICS

This program will build a recommendation engine for lyrics in Spanish using Artificial Intelligence and Natural Language Processing.

Project led by Elena González-Blanco

www.lyraics.com
Conclusion

The application of AI&NLP to Spanish music will make our algorithms more “human”, but it will arrive much further…
Chatbots

Medical records
Fintech, banking and insurance

charles Schwab
FutureAdvisor
sift science
Betterment
Kensho
zest finance
motif
feedzai
Personal Capital
wealthfront
Nutmeg
Energy
Connected car

Legal documentation
El español. El segundo idioma más hablado del mundo
470 millones de hablantes

Challenges
Challenges

El español, el segundo idioma más hablado del mundo a 470 millones de hablantes
Challenges
Challenges
Challenges
“Will robots inherit the earth? Yes, but they will be our children”.
Marvin Minsky, Scientific American (October 1994)
Thanks!

Any questions?

You can find me at:
@elenagbg &
egonzalezblanco@faculty.ie.edu