**Project proposal**

*Answer these questions concisely below each other according to the given numbering. Use the Cambria font, no Bold.*

**Date: 23 March 2020**

**Name of students:**

**Tsz Kin Chau tszkin.chau@student.kuleuven.be**

**Fengan Li fengan.li@student.kuleuven.be**

**Hugo Enrique Montaño Castillo hugoenrique.montanocastillo@student.kuleuven.be**

**Lars Magne Tungland larsmagne.tungland@student.kuleuven.be**

**Martina Verna martina.verna@student.kuleuven.be**

**Xin Wang xin.wang@student.kuleuven.be**

1. **Motivation.**

What is the problem that the visualization will try to explain or solve? Why is this problem important? Why is this problem complex?

Problem: 1) To show our capacity to deal with big datasets; 2) Music is important pop culture. The lyrics haven’t been analyzed in detail. The vocabularies

Why the problem: How music industry came into being: assumption: songs need to be fit in the higher ranking not only thanks to the melodies but also the content/message it tries to convey.

Why is this problem complex: it’s a big dataset to begin with, plus its particular way of constructing the words; large and makes of words; it’s complex to manipulate large amount of texts

1. **Insight.**

What data-driven insight(s) do you expect your visualization to be able to reveal?

(Be aware that the visualization will have to be developed primarily for this objective)  
  
Insight 1: the change of the frequency of the words in relation to the popularity of songs

Insight 2: How the frequency of words changes by genre and period

Insight 3:

1. **Data.**

Which datasets will you use? What kind of data attributes do they contain? What kind of licenses are connected to these datasets? How big are they? How recent are they? How often are they refreshed? Etc.

(Fill in the table below, or make your own table)

Title dataset 1:

Source:

Data attributes:

License:

...

1. **Inspiration.**

Which existing visualizations will inspire you? Which aspects specifically do you find interesting? Why?

(Fill in the table below, or create your own table; Give at least 3 different examples)

Title example 1:

Short description:

Source:

Inspiration aspect + argumentation:

Relevant screenshots (copy / paste image (s)):

...

1. **Provisional sketches**

What will your visualization potentially look like? What are the interactive possibilities? What will be the narrative with which the user can use the visualization and understand the insights?

(Sketch - by hand, Photoshop or through other simple means - some strategies that illustrate how the user will come to the insights of point 2 via the visual decoding, interactive functions and narrative storytelling of the visualization) (if there are multiple ideas, display them below)

Idea 1

Screenshots (copy / paste image (s)):

Short description:

Name visualization technique:

Interaction:

Narrative:

...

1. **Questions / problems**

Data cleaning: null values

The categories in two genre datasets are either too general or too specific

Different languages of the songs cannot be treated separately

Biased in the datasets in relation to the period the songs are collected

Data stop at about 2010