

Automated analysis of mission statements to group arts organisations by their activity focus and communities served

Methods to learn and understand particularities of plain text data without human help

This project was developed using data gathered from http://artbridges.ca/community/community_map. The dataset contains a small number of mission statements submitted by community art centers in Canada, as well as the associated target community served. To increase the size of the original dataset, data on international community art centers were also collected from <https://www.icasa.ca/directory>.

Two Goals: Automatically group arts community centers into semantically similar categories and find the target clientele for each establishment based on their mission statement.

Limitations:

This task is challenging primarily due to the small number of statements. Modern language processing techniques require a lot of data to capture semantic idiosyncrasies. We addressed this problem by introducing additional data - with a modest number of domain-specific texts, combined with a large amount of generic text. Automatic assignment of target clientele is not sensible due to the way the data was entered - a limited number of specific, non-overlapping categories would be ideal. Instead, we built a model to associate different texts together, matching the type of clientele with each mission statement.

Grouping Arts Centers: One challenge is to determine the ideal number of categories. Four categories consistently yielded the best results. This number could be increased with additional data. Below is a sample of keywords selected in each category. Note that some categories share keywords, as meaning differs depending on context.

Four Topics Automatically Distributed From the ASC Dataset

Topics	Work	Performance Arts	Education	Social Issues
Keywords	program work workshop project participate collaboration experience	theatre dance creativity performance audience story express	educate youth young people opportunity public develop mission	awareness world politics HIV women support space

Categorizing Clientele: The model was able to detect words related to the target clientele. This is a good starting point to classify the arts centers by community served without human help. The plots below are a small sample of randomly selected words and phrases. The different colours represent the machine understanding of semantic groups in the text.

