UNLP at the MediaEval 2014 C@merata Task

Kartik Asooja
Insight Centre for Data Analytics,
NUI Galway, Ireland
kartik.asooja@insight-centre.org

Sindhu Kiranmai Ernala
Center for Exact Humanities,
IIIT Hyderabad, India
sindhukiranmai.ernala@students.iiit.ac.in

Paul Buitelaar
Insight Centre for Data Analytics,
NUI Galway, Ireland
paul.buitelaar@insight-centre.org

ABSTRACT

This paper presents a description of our submission to the C@merata task in MediaEval 2014. The system answers the natural language queries over the musical scores. The approach is based upon two main steps: identifying the musical entities and relations present in the query, and retrieving the relevant music passages containing those entities from the associated MusicXML file. We submitted two runs for the task. The first one takes a union of the passages retrieved for each musical entity, while the second approach takes their intersection to answer the query. Musical entities in the query are recognized with the help of regular expressions.

1. INTRODUCTION

This work explains our system submitted in the C@merata task [1] at MediaEval 2014. The task targets natural language question answering over the musical scores. We were provided with a set of question types, and the data over which the search was required to be performed.

The questions in the task consist of short noun phrases in English referring to musical features in the music scores, for instance, "F# followed two crotchets later by a G". Every question refers to a single natural language noun phrase using English or American music terminology. The music scores are provided in MusicXML [2], which is a standard open format for exchanging digital sheet music. The music repertoire consists of Western Classical works from the Renaissance and the Baroque periods by composers like Dowland, Bach, Handel, and Scarlatti. The answers comprise of the music passages from the music score containing the musical features mentioned in the query string. Thus, it points to the location(s) of the requested musical features in the score. The answer passage consists of start/end time signature, start/end division value, and start/end beat. The task provides two datasets, one for development consisting of 36 natural language queries while the other for testing containing 200 questions.

2. APPROACH

There can be different types of musical features mentioned in the query such as note, melodic phrase and others. These different musical features can be referred as musical entities or can be defined with the help of such entities. Therefore, we identify some of the basic entities from the natural language text, and perform the location search by comparing the extracted entity values against the corresponding values in the music score for retrieving the answer passages. In the current implementation, we recognize only basic musical entities. For the complex ones requiring some combinations according to particular relations between the entities, we just take the union or intersection of the answer measures retrieved separately for different entities appearing in the query. Thus, our approach consists of the following two main steps: Identification of musical entities in the query, and retrieval of the relevant music passages from the provided MusicXML file. Figure 1 summarizes the followed approach.

Copyright is held by the author/owner(s).
MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain.

2.1 Identification of Musical Entities

We use regular expressions and created dictionaries to recognize musical entities in the query strings. The target entity types are:

1. Notes: A note defines a particular pitch, duration or dynamic, such as C, crotchet C, quarter note C in the right hand, semibreve C. The note recognizer comprises of three basic music entity recognizers: duration, pitch and staff. We first recognize all the pitches appearing in the query string, and separately identify all the durations and staves. To assign the correct duration/staff for a pitch, we measure the string distance between all the pitches and duration/staff. The duration/staff, which occurs within a threshold distance from a pitch, is paired with it in order to form the note. The pitches and durations are identified using regular expressions.

Duration: It defines the playing time of the pitch. In natural language, it can be reflected by the terms like quarter, semibreve, and whole. We write a regular expression covering the extensive vocabulary defining the duration in both English and American music terminology.

Pitch: It is a perceptual property that allows the ordering of sounds on a frequency-related scale. Some examples of writing pitches in natural language are: D sharp, E#, and A flat. We form a regular expression to identify the pitches in a query string.

Staff: To identify the staves mentioned in a string, we find the occurrences of "right hand" and "left hand" strings in it.

The three basic musical entities: duration, pitch and staff collectively form the note entity.

2. Instruments: In order to find the instruments mentioned in the query string, we manually created a dictionary of instrument related n-grams using the training and test data. The dictionary includes the words like viola, piano, alto, violoncello, soprano, tenor, bass, violin, guitar, sopran, alt, violin, voice, and harpsichord.

3. CLEF: To identify the Clef, we just check the presence of strings like bass clef, F-clef, treble clef and G-clef in the query.

The implementation including the regular expressions and the dictionaries used can be found at the publicly available code repository at GitHub[1].

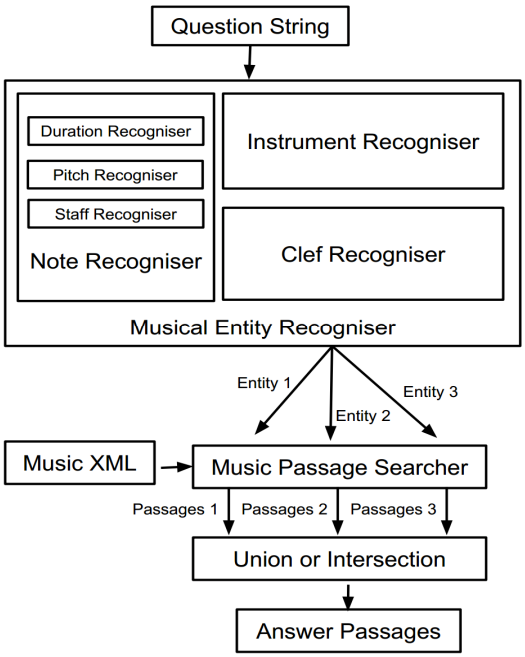

Figure 1: Approach

2.2 Music Passage Retrieval

The values of the identified musical entities in the query are compared against the corresponding values extracted from the music score xml file associated with the question. The identification of the musical entities remains same in both the submitted runs. They just vary on the basis of the following two approaches for music passage retrieval:

1. The union of the musical measures that contain the target musical entities is used to create the answer passages.

2. An intersection of the musical measures that contain the target musical entities is used in the answer passages.

3. RESULTS AND DISCUSSION

The system performance is measured for each question type, and an overall weighted average for all the questions is also calculated. Table 1 shows the results obtained by our two runs. As discussed in the approach section, the current implementation recognizes only a few types of musical entities, which constraints the question types to be answered. The results clearly show that the system could not answer many question types like melodic, harmonic, and cadence. It is because the system could not detect such musical features.

Our system only uses regular expression matching for the identification of musical elements, and string distance for to identify the relations between the elements wherever required. However, there is a scope of deep syntactic and lexical analysis of the query string to better identify the relations between the entities. We also found a minor bug in our system related to the "natural" appearing in a query string. It led to some wrong answers because of incorrect octave calculation, which is now updated in the current implementation at GitHub.

The second run gives a much better measure precision than the first, especially in the question type "Followed by". It is because such queries contain different notes separated by "Followed by", and the union approach merges all the measures retrieved from the notes decreasing the precision, while the intersection just gives those measures, which contain both the notes. However, given query types other than "Followed by" do not generally contain more than one type of notes, therefore, similar scores are generated for both the runs.

Table 1. Result table

Query Type	Beat Precision Run 1|2	Beat Recall Run 1|2	Measure Precision Run 1|2	Measure Recall Run 1|2
Overall	0.11|0.29	0.52|0.51	0.16|0.39	0.70|0.69
Pitch	0.42|0.42	0.79|0.79	0.48|0.48	0.89|0.89
Length	0.64|0.64	0.80|0.80	0.79|0.79	0.99|0.99
Pitch & Length	0.46|0.46	0.70|0.70	0.58|0.58	0.88|0.88
Perf.	0.05|0.05	0.59|0.59	0.05|0.05	0.69|0.69
Stave	0.17|0.17	0.44|0.37	0.23|0.24	0.59|0.52
Word	0.07|0.07	0.83|0.83	0.07|0.07	0.83|0.83
Follow-ed by	0.00|0.00	0.00|0.00	0.03|0.26	0.70|0.63
Melodic	0.00|0.00	0.00|0.00	0.00|0.00	0.00|0.00
Harm-onic	0.00|0.00	0.00|0.00	0.00|0.00	0.00|0.00
Cadence	0.00|0.00	0.00|0.00	0.00|0.00	0.00|0.00
Triad	0.00|0.00	0.00|0.00	0.00|0.00	0.00|0.00
Texture	0.00|0.00	0.00|0.00	0.00|0.00	0.00|0.00

4. CONCLUSION

The proposed approach presents an initial implementation of the natural language question answering on musical scores. The pipeline is based upon identifying the different types of musical entities and their relations in the query string, and comparing them against the corresponding values extracted from the MusicXML file to identify the answer passages. We consider applying natural language processing on the queries to better extract the music entities and relations, as a future direction to explore.

5. ACKNOWLEDGEMENTS

This work has been funded in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (INSIGHT) and by the EU FP7 program in the context of the project LIDER (610782). We are very grateful to Mr. Robert Solyom from Music Academy, Galway for helpful suggestions and references.

6. REFERENCES

[1] R. Sutcliffe, T. Crawford, C. Fox, D.L. Root and E. Hovy. The C@merata Task at MediaEval 2014: Natural language queries on classical music scores. In MediaEval 2014 Workshop, Barcelona, Spain, October 16-17, 2014.

[2] M. Good, and L. L. C. Recordare. Lessons from the Adoption of MusicXML as an Interchange Standard. 2006.

[1] https://github.com/kasooja/camerata