UNLP at the MediaEval 2015 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 2015. This submission is a revision to the system submitted for the same task in MediaEval 2014 including some bug fixing. 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. Our approach makes a sequence of the musical entities in the query, and then searches for a sequence of passages satisfying the sequence of the entities. 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 2015. The task targets natural language question answering over 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][3], 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 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 training and development consisting of the 236 natural language queries used for the task last year while the other dataset is newly introduced for testing, which contain 200 questions. This year, questions are linguistically more difficult and the scores are more complex.

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 to as musical entities or can be defined with the help of such entities. Therefore, we identify some of the musical entities from the natural language text, and perform a location search by comparing the extracted entity values against the corresponding values in the music score for retrieving the answer passages. For the complex queries requiring some combinations according to particular relations between the entities, we just consider the sequential relation between the musical entities as they appear in the query string. Rather than making a system,which differentiates between question types, we apply a rather simple approach assuming just the sequential relation. On the other hand, the approach we submitted last year performed union or intersection of the answer passages found for each musical entity. Our current approach consists of the following two main steps: Identification of the sequence of the musical entities in the query string, and retrieval of the answer sequences of the relevant music passages matching the sequence of entities. Figure 1 summarizes the followed approach.

MediaEval 2015 Workshop, September 14-15, 2015, Wurzen, Germany.

2.1 Identification of Musical Entities

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

1. Notes: A note defines a particular pitch, duration or dynamics using strings such as Do, crotchet C, quarter note C in the right hand, or 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: This 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: This 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 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].

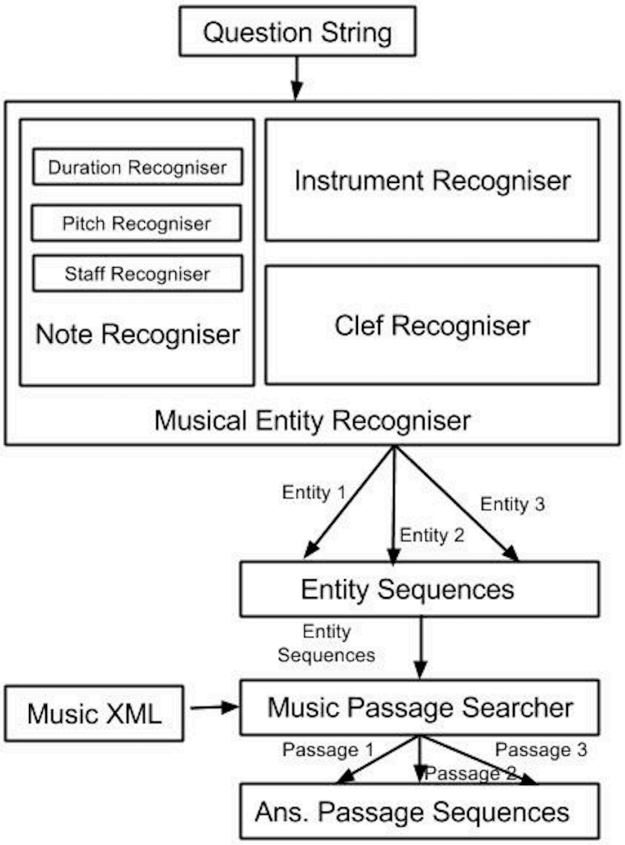

Figure 1: Approach

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 system matches the musical features sequentially as they appear in the query string. Finally, the passage sequences matching completely with the sequence of musical entities are selected as 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 submission for some question types. As discussed in the approach section, the current implementation only recognizes a few types of musical entities, which constrains the question types to be answered. The results clearly show that the system could not answer many question types like texture, harmonic etc. It is because detection of such musical features was not implemented in the current system. Comparing to the system submitted last year, we removed many bugs related to the meaning of different tags in the Music XML reader, as we implemented our own reader in Java. In the current version, our system only uses string and regular expression matching for the identification of musical elements, while string distance is used to identify the relations between the elements, if required. However, deep syntactic and lexical analysis of the query has the potential to identify relations between the entities more accurately.

Table 1. Result table

Query Type	Beat Precision	Beat Recall	Measure Precision	Measure Recall
Overall	0.126	0.43	0.149	0.508
1 Harmonic	0.0	0.0	0.0	0.0
Synch	0.0181	0.194	0.0207	0.222
1 Melody Alone	0.79	0.942	0.79	0.942
Perf.	0.0789	0.6	0.0877	0.667
Instru.	0.562	0.202	0.562	0.202
Clef	0.145	0.481	0.157	0.519
Followed by	0.25	0.0968	0.625	0.242
1 Melody	0.406	0.769	0.408	0.773
n Melody	0.0247	0.196	0.058	0.461
Key	0.0	0.0	0.0	0.0
Time	0.208	0.762	0.247	0.905
Texture	0.0	0.0	0.0	0.0
1 Melody Clef	0.875	0.875	0.875	0.875

4. CONCLUSION

We have presented a simple pipeline for 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.

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). 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 2015: Natural language queries on classical music scores. In MediaEval 2015 Workshop, Wurzen, Germany, September 15-16, 2015.

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

[3] Music XML: http://www.musicxml.com/

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