Addendum to "Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics"

Technical Report

Peter Knees
Department of Computational Perception
Johannes Kepler University Linz, Austria
peter.knees@jku.at

ABSTRACT
The purpose of this technical report is to discuss two additional aspects of automatic lyrics retrieval as described in "Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics" by Knees et al., 2005. The first aspect is the introduction of a confidence measure to estimate the quality of the generated output. The second aspect deals with the automatic formatting of generated lyrics to present the user with results that look like manually produced lyrics.

1. INTRODUCTION
This paper builds upon the techniques presented in [1] to automatically retrieve correct versions of song lyrics from the Internet. Two aspects (that could not be included in [1] mainly due to space restrictions) are presented and discussed: a *confidence measure* to estimate the quality of the generated output and an approach to *automatically formatting* of the generated lyrics to give the impression of manually produced lyrics. Note that before reading this report, the reader should be familiar with the notations, terms and techniques introduced in [1] since this report is only intended as an addendum to the original paper.

2. CONFIDENCE ESTIMATION
The approaches presented in [1] aim at generating an output that is (as much as possible) correct by incorporating different sources. Besides the mere generation of such a consensus-based result, it is also desirable to have a confidence measure to estimate the quality of the output. This additional information can be valuable to reject results if certain quality requirements are given. Estimation of the quality also allows the system to automatically point out problems with certain lyrics to the user, for example in a batch tagging scenario with a large number of songs. Furthermore, a low confidence might be used to trigger another query with different keywords. For example, the lyrics generated from pages retrieved by the query `"Shakira" "Que Me Quedes Tú" lyrics` might produce a result with low confidence. As a consequence, the system could automatically generate lyrics via the query `"Shakira" "Que Me Quedes Tú" letras` and present this version if its confidence value is higher.

Technical Report CPJKU-TR-2008-MLA.

We propose a heuristic that incorporates both the certainty of decision for the single words and the coherence of the output string. Therefore, we estimate the quality of the presented output as follows:

$$conf(l) = \frac{len(\widetilde{l})}{len(l)} \cdot \frac{1}{len(l)} \sum_{i=1}^{len(l)} cert(l_i) \tag{1}$$

where $\widetilde{l}$ is the smoothed version of l, l_i the i^{th} word in l, $len(x)$ the length of x,

$$cert(x) = \frac{maxword(x)}{depth(x)}, \tag{2}$$

and $maxword(x)$ the number of occurrences of the most frequent word in x. Using the ratio $\frac{len(\widetilde{l})}{len(l)}$ we assess the "stability" of the result. For sequences containing the lyrics with only a few scattered words around, the ratio is nearly 1 and thus the confidence is estimated primarily by the average agreement of the words. For alignments built accidentally on pages not containing any lyrics, but any other text, the confidence is very low, since the alignment will consist mainly of incidentally matching stopwords, which will be omitted during smoothing.

To evaluate the expressiveness of the confidence measure for prediction of result quality, we computed the correlation of confidence and recall for all values of t on the same lyrics test corpus as in [1]. The result is visualized in Figure 1. First of all, it can be seen that the confidence correlates much more strongly with results achieved with the tfcorr page selection approach (especially with the smoothed version), than with those achieved with the kwit approach. A remarkable fact is the (partly) high correlation for high values of t. For tfcorr this could be explained by both decreasing recall and confidence (cf. Figure 2). However, for kwit this consideration does not hold. Since comparing the means of two distributions reveals nothing about the correlation between them, these conclusions are not satisfying. Rather, the reason might be the fact that confidence tends to underestimate the actual recall. This tendency vanishes with growing t, leading to higher correlation. This observation can be made for both approaches, whether smoothed or unsmoothed. The fluctuations at the right end are probably not significant and thus only indications of the volatility of the approach for high values of t. This is discussed in more detail at the end of this section.

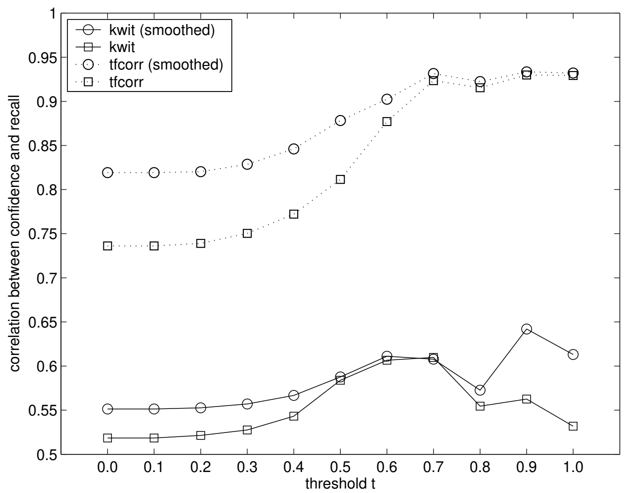

Figure 1: Correlation of confidence and recall for both page selection approaches (smoothed and unsmoothed).

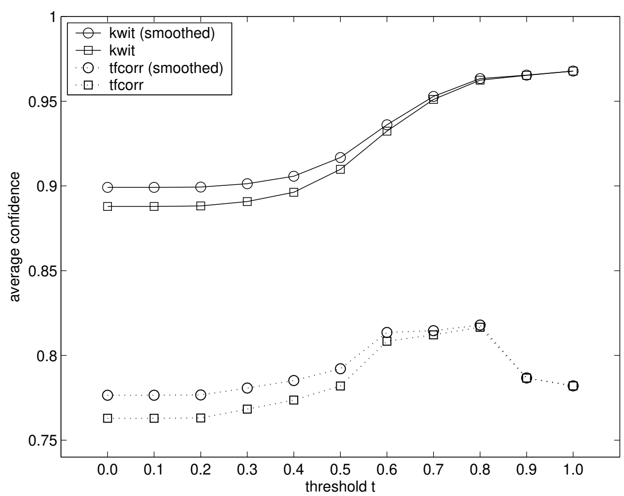

Figure 2: Average confidence measure for both page selection approaches (smoothed and unsmoothed).

For our preferred t values of 0.5 and 0.6, the correlations of confidence and recall for the unsmoothed tfcorr approach are 0.81 and 0.87, respectively. For the smoothed version, correlation is even higher with values of 0.87 and 0.9. For the kwit approach correlation is around 0.6 for both t=0.5 and t=0.6.

Finally, we want to examine the development of the confidence measure and find a reasonable explanation for the unexpected results in Figure 1. Since for t=1.0 the average certainty in Equation 1 is 1, the confidence is determined solely by the ratio $\frac{len(\tilde{l})}{len(l)}$. For kwit, this ratio is high, indicating little loss of words by the use of smoothing and thus only slight word fragmentation in the result. For tfcorr this does not seem to be the case. The quality of underlying pages is likely to be the important difference between the approaches.

3. AUTOMATIC LYRICS FORMATTING

The second goal of this paper is to advance the original lyrics alignment method by appending a formatting step to output easily readable results that look like manually produced lyrics. To format the lyrics, we basically reuse the techniques from [1]. After finding the most probable sequence of words, in a second step we again apply Sequence Alignment to reobtain structural information discarded throughout the lyrics alignment process. To this end, we perform pairwise alignment between the result string (i.e. the output produced by the method in [1]) and all of the original sequences (i.e. the underlying web pages) to find the sequence most similar to the result. Our assumption is that the most similar sequence is best suited to provide a structure for the output.

To transfer the structure to the result, we modify the alignment procedure to preserve line breaks, dots, colons, commas, question marks, and exclamation marks. Furthermore, we have to adapt the scoring scheme of the Sequence Alignment to ensure that none of these special characters are aligned to any of the words from the result. The third modification concerns the preprocessing of the web pages. While in the original approach all characters have been converted to lower case to simplify the alignment, we preserve upper case letters. This has no influence on the alignment, since comparison of strings is performed on simplified versions (for details see [1]).

From the alignment of the result and the best matching sequence, we adopt all words from the result sequence and all special and control characters from the best matching sequence. The next step is to clean up the output, since it usually contains a lot of unnecessary punctuation. This originates from those parts of the web page that are unrelated to the lyrics. Thus, we remove all punctuation and line breaks that occur before the first word and after the first punctuation mark after the last word. Furthermore, for successive punctuation marks we keep only the first one and allow only a maximum of two successive line breaks.

The formatted result of a query for the song *Yesterday* by *The Beatles* can be found in Figure 3. Beside the fact that the retrieved content is perfectly accurate, our formatting technique produced an output that looks like manually typed lyrics. However, obviously the first line has been detached from the remaining lyrics by insertion of an empty line. This is caused by the fact that the song's first word is not aligned to the first word of the lyrics on the web page but to the title of the song which precedes the actual lyrics. Accepting such small variations, our formatting method is an useful extension to the automatic lyrics retrieval method from [1]. Giving the impression of lyrics edited by a human, although the content is automatically merged from different sources across the web, the approach can serve as an "intelligent" replacement for existing lyrics download applications.

4. CONCLUSIONS

We have presented two enhancements to the Multiple Lyrics Alignment approach presented in [1]: first, a confidence measure to estimate the quality of the generated output, and second, a technique to format the output of the original lyrics retrieval approach.

The confidence measure showed to be a good indicator for the quality of the generated lyrics. By calculating this additional information, we enable advanced retrieval approaches that allow for dynamically adapting the query settings (e.g. for lyrics in languages other that English). Furthermore, in semi-automatic retrieval tasks, the user's attention could be directly drawn to difficult cases.

Using the formatting method, we generate an easily readable output that gives the impression of manually typed lyrics instead of an unstructured sequence of words. To this end, we transfer the structure of the most similar web page to the unstructured sequence. Since there exist numerous alternatives to format lyrics, a quantitative evaluation of the results is difficult. However, by incorporating this extension, the approach is directly applicable as automatic lyrics download tool (e.g. incorporated into music player software), making it also interesting to the average user.

5. ACKNOWLEDGMENTS

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project numbers L112-N04 and L511-N15.

6. REFERENCES

[1] P. Knees, M. Schedl, and G. Widmer. Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics. In *Proceedings of 6th International Conference on Music Information Retrieval (ISMIR'05)*, pages 564–569, London, UK, September 2005.

Yesterday

All my troubles seemed so far away,
Now it looks as though they're here to stay,
Oh, I believe in yesterday.
Suddenly,
I'm not half the man I used to be,
There's a shadow hanging over me,
Oh, yesterday came suddenly.
Why she
Had to go I don't know, she wouldn't say.
I said,
Something wrong, now I long for yesterday.
Yesterday,
Love was such an easy game to play,
Now I need a place to hide away,
Oh, I believe in yesterday.
Why she
Had to go I don't know, she wouldn't say.
I said,
Something wrong, now I long for yesterday.
Yesterday,
Love was such an easy game to play,
Now I need a place to hide away,
Oh, I believe in yesterday.

Figure 3: Formatted output of automatic lyrics retrieval for *Yesterday* by *The Beatles*.