Using Speech Recognition and Intelligent Search Tools to Enhance Information Accessibility

Keith Bain[1], Jason Hines[2], Pawan Lingras[2]

[1]Liberated Learning Consortium, Saint Mary's University, Halifax, NS B3H 3C3, Canada,
keith.bain@stmarys.ca
[2]Mathematics and Computer Science Department, Saint Mary's University, Halifax, NS
B3H 3C3, Canada, j_hines@cs.smu.ca

Abstract. Access to spoken information presents challenges in educational, work, and public settings for many individuals. Speech Recognition technology offers a potential strategy for increasing access in various settings, including the university lecture environment. An international research team has successfully experimented with providing speech recognition generated lecture notes, available online as multimedia enhanced transcripts. Additionally, a specially designed search and indexing tool was introduced to enhance the user experience. An exploratory study was conducted to gauge student reactions, identify usage patterns, and assess ease of use.

Keywords: speech recognition, transcription, multimedia, search, retrieval, indexing

1 Introduction

Although physical access has improved in educational, employment, and public environments, information accessibility remains a significant barrier for many individuals. Information accessibility challenges abound in everyday situations. Blind and visually impaired persons face challenges accessing print information without intervening assistive technologies. Standard electronic documents can likewise introduce barriers. For example, early PDF documents were largely inaccessible for visually impaired users. Without available captions, digital audio is similarly inaccessible for those who are Deaf and/or Hard of Hearing. Other stakeholders, such as English Second Language learners, often struggle with audio information delivered in non-native languages.

Efforts to improve access to spoken information have garnered increased attention. For example, the Liberated Learning Consortium (www.liberatedlearing.com), comprised of fifteen member institutions in eight countries, is researching and

developing new speech recognition (SR) technologies that improve information accessibility. Using SR to caption real-time speech and generate multimedia enhanced transcripts is subsequently referred to as the Liberated Learning concept. The initial driver for Liberated Learning was dissatisfaction with conventional, intermediary based note taking approaches for students who were unable to write their own notes due to disabilities. Targeted stakeholders initially included those with physical and hearing disabilities, given these populations were most likely to require note taking support. The earliest studies considered the viability of using SR as live captioning and transcription tool for creating class notes in a university lecture. For these early software generated "notes", recognition errors were corrected, punctuation was added, formatting applied, summarization considered, and in some instances the prose itself was improved. These interventions were deemed necessary to transform the raw transcripts into a form more aligned with traditional definitions of class notes. These activities added significant overhead to the entire concept workflow and were generally acknowledged to be unsustainable given available resources. A gradual paradigm shift moved researchers away from the notion of providing SR based notes to the provision of SR generated transcripts. Activity therefore shifted from document enhancements to exploring various data alternatives, strategies for integrating archival and distribution platforms, and most importantly, enhancing user experience.

Two principal challenges emerged. Early users complained about the inherent sequential access nature of the text based transcripts, which were provided in hard copy for interested students. Because typical sixty minute lectures contained between 7,500 and 10,000 words, printing costs were prohibitive, as well as logistically difficult to manage. Transcripts were gradually made available through an online notes system. However, the majority of students downloaded the text documents to their local machines, where they typically assumed the burden of printing the entire lecture. Although students responded positively to the availability of SR generated transcripts, they reported an interest in accessing a variety of data formats which suited their individual learning preferences. They also expressed interest in the availability of explicit tools that would facilitate more efficient and intelligent utilization of the transcripts [1].

Two simultaneous efforts emerged to address these issues. First, developers designed a new SR application called IBM ViaScribe, which among other innovations, produced multimedia-based transcripts [2]. Using the W3C standard, developers published Synchronized Multimedia Integration Language (SMIL) based outputs. SMIL provided a standardized method for maintaining the synchronization between the resulting text transcription and the instructor's original audio, while in theory allowing students to access content through freely available media players. Researchers also published data using Microsoft's Synchronized Accessible Media Interchange (SAMI), given Window's Media Player did not conform to W3C standards. Eventually, multimedia output was integrated directly into an html structure.

Secondly, researchers began to conceptualize ways to improve machine readability of these multimedia transcripts and develop intelligent search and retrieval tools that could improve the user experience. A key premise was that students did not want to necessarily listen to/read/review the lecture in its entirety, but rather have control over

the transcript experience. Researchers believed that students would opt to self select specific sections, search for key words, or extract concepts that were individually important for subsequent study. Furthermore, faculty users could potentially benefit from tools that helped them revisit particular lectures or index keywords from a corpus of SR generated course transcripts.

2 Search and Retrieval of Keyphrases

Early information retrieval systems used Boolean logic to search for existence of the given keywords in a document [3]. Salton proposed the use of relative weights of terms in what he termed as the vector space model. In the conventional vector space method [4], the document collection is represented by an $m \times n$ term-document matrix, where m is the number of terms and n is the number of documents. Typically this matrix has fewer than 1% nonzero entries. Queries are represented as m-vectors, and a matrix-vector product produces an n-vector of scores that is used to rank the documents in relevance.

For convenience, initially, the vector space model (VSM) used automatically extracted keywords for identifying relevant documents. However, VSM can just as easily be applied to keyphrases. Previous research has shown that using keyphrases to assist the information retrieval processes can increase the retrieval efficiency.

Fagan [5] and Mitra [6] show that phrase-based indexing helps to increase the precision of the overall retrieval, especially at lower relevance ratings, even though it does not significantly improve the result at higher relevance ratings where phrases tend to overemphasize a particular aspect of the query. Anick [7] developed a relevance feedback system Paraphrase based on the observation that "lexical dispersion" of a term predicts the likelihood of the term being a topical term. The lexical dispersion is defined as the number of different noun compounds containing that term appearing in a given document set. Anick contended that noun compounds or keyphrases provide more discriminating power than individual terms. Phrases were extracted from the initial result set and presented to the user for relevance feedback. Another such attempt in the form of a document retrieval system called Phrasier was reported by Jones [8]. Phrasier ranked the similarity between documents by the degree of keyphrases overlapping in the documents. Here the query is also treated as a document. Results from human evaluation of Phrasier indicate that the phrase-based retrieval system is as effective as a full-text retrieval system while requiring less storage space for indexing. Buckland et al. [9] use noun phrases in the titles, authors and/or abstracts and metadata vocabularies such as classifications to build an "Entry Vocabulary Module" (EVM). EVM associates the noun phrases and the metadata vocabularies with the likelihood ratio statistics. The results are used as a dictionary for translating user queries in a natural language, such as English, to corresponding metadata vocabularies that are unfamiliar to the user. Since utilizing phrases is beneficial in many aspects in information retrieval, finding a set of keyphrases for a document repository that is suitable for both indexing and document retrieving is essential to improving efficiency of retrieval systems. Keyphrases also play an important role in exchanging and retrieving information with multiple repositories. In

addition, keyphrases can help users get a better understanding of the content of a collection. They provide sensible entry points for the collection, show how queries can be extended, facilitate document skimming by visually emphasizing important phrases, and offer a powerful means of measuring document similarity [10] [11].

Traditionally, keyphrases are chosen manually. Authors assign keyphrases to documents they have written. Professional indexers often choose phrases from a predefined *controlled vocabulary* relevant to the domain at hand [12]. However, most of the documents on the web do not have keyphrases associated with them. Manual assignment of keyphrases is a tedious process that requires knowledge of the subject matter. Therefore, researchers are developing automatic extraction techniques.

There are two fundamentally different approaches to keyphrase generation: keyphrase assignment and keyphrase extraction [13]. Both use machine learning methods, and require a set of documents with associated keyphrases for training. Witten, *et al* [14] proposed Kea, keyphrase extraction algorithm. Kea generates candidate phrases by looking through the input document for any sequence of one, two, or three consecutive words. The consecutive words must not be separated by punctuation and must not begin or end with stop words. Stop words are commonly occurring words such as "the", "of", "to", "and", "he". Candidate phrases are normalized by converting them to lower case and stemming them. Stemming is a process of linguistic normalisation, in which the variant forms of a word are reduced to a common form, for example, the words: "connection", "connections", "connective", "connected", and "connecting" are all stemmed to "connect". Kea then uses the naive Bayes algorithm to learn to classify the candidate phrases as either keyphrase or nonkeyphrase. Turney [14] provided evidence that statistical word association can be used to improve the coherence of keyphrase extraction, resulting in higher quality keyphrases, measured by the degree of overlap with the authors' keyphrases. Furthermore, the new coherence features are not domain-specific. This work has been further extended to extraction of keyphrases from spoken language documents by Inkpen and Desilets [15] [16].

3   Note Finder Personal Search Engine

Utilizing the above keyword/keyphrase generation and extraction approaches, a prototype search and retrieval tool, named "Note Finder" was developed in 2005. Note Finder provides searching abilities based on keyphrases and stemming, automatic phrase table generation, listing of all the course documents, and the ability to crawl and index external websites containing various file types. The system updates automatically on a scheduled basis. Note Finder is built on the open source LAMP (Linux, Apache, MySQL, and PHP) framework which makes the source code accessible.

Note Finder works like Kea, with the exception that a phrase is always two words. Each indexed word has two lists associated with it; a same-stem list, consisting of words which have the same stem, and a proximity list, consisting of words which appear within a threshold distance of +/- five words (and do not contain any stop words).

The same-stem list is useful for displaying the various forms of a word in the phrase table. The proximity list consists of quadruplets (keyword, frequency, average distance, and minimum distance) which are used to determine the relevance of keyphrases. Using these two lists, a software agent automatically generates a phrase table (Fig. 1) each time the system is updated.

Fig. 1. Note Finder Phrase Table

Unlike conventional search engines, users have access to the top one hundred keyphrases. The phrase table gives users the ability to see the keyphrases which exist within the document set and also provides hyperlinks to quickly search for those keyphrases. This feature is especially useful for aiding memory recall. For example, users may be looking for a unique phrase but may not remember the exact semantics of the phrase. Since the phrase table presents users with keyphrases and a list of similarly spelled words, there is a greater chance that the user will find what they are looking for.

Fig. 1 shows the layout of the phrase table. From left to right, users are first presented with a collapsible menu of root keywords, a 'Similarly Spelled' link to display words which are similar to that keyword (contain the same stem), a 'Note Finder' link to search for that keyword in the Note Finder system, and a 'Google' link to search for that keyword using Google. By expanding the root keywords in the collapsible menu, users are presented with a sub-list of keywords. Each of the keywords in the sub-list belongs to a phrase containing the root keyword. As shown in Fig. 1, "latin america" and "north america" are keyphrases. Users can then search

for these keyphrases by clicking on the corresponding 'Note Finder' or 'Google' link. The phrase table is sorted in alphabetical order and only displays ten keyphrases per page. To make the phrase table easier to read, different background shading is used for each alternating set of keyphrases.

Note Finder also provides a section for users to retrieve notes manually. This is useful for instances where users know exactly what document they are looking for. This is all done in the 'Indexed Files' page of the system and is presented similar to that of the phrase table with the exception of using a collapsible directory structure.

Search results are quite similar to that of any major search engine but again Note Finder makes use of background shading to clearly differentiate between results. The document title is presented as a hyperlink through which the user accesses the current document. In instances where the document title is non-existent the actual filename is presented. If the document is a Liberated Learning enhanced multimedia document then users are presented with a link to the standard text format of the document as well as a link to the multimedia enhanced version of the document. Users also have the option to view a cached version of the document in which a list of same-stem search terms will be automatically highlighted.

5   Evaluating SR Multimedia Transcripts and Note Finder

In the 2006 fall semester, the first full term usage of Note Finder occurred. Since the system seemed robust from a technical and administrative perspective, the project team wanted to assess whether students found the system useful. Two, introductory undergraduate courses in Sociology and Anthropology were selected for evaluation. Both professors were experienced users of SR, having utilized the "Liberated Learning" approach for a number of years. The courses were also largely lecture based with large enrolments. In most classes, the professors used IBM ViaScribe to transcribe their lectures. The resulting multimedia transcriptions were saved to an automated network archival system [17]. A third party researcher subsequently uploaded the files to a dedicated server that hosted the Note Finder system. The Sociology professor additionally uploaded the multimedia transcripts to an internal network drive, which contained other course materials (syllabus, readings, etc). This drive was only accessible while physically on campus by logging into the university's local area network. The Anthropology lectures were only available through the Note Finder interface, which was accessible over any standard browser. The digitized lectures were typically available by the end of the day.

To help ensure feedback was as unbiased as possible, the researchers attempted to reinforce that system usage was not tied to course evaluation. At the beginning of the semester, an independent researcher provided both classes with an overview of the Liberated Learning concept and explicit instructions on accessing and using Note Finder. It was conveyed that the professors would not have system access and therefore would not be privy to usage statistics. Similarly, students were informed that researchers would not analyze usage on an individual basis. Throughout the term, both professors indicated they did intermittently remind students of the availability of

the multimedia notes and Note Finder tool, but did not attempt to gather anecdotal feedback or discern if particular students were using the system.

For evaluation, a simple survey (Fig. 2) was developed and administered to gauge student reactions. There was an assumed implicit linkage between the SR generated transcripts and the tool students would use to access these transcripts. In other words, measuring reactions to the Note Finder tool itself would provide insights into the perceived underlying value of SR generated multimedia transcripts. There were no specific questions about the SR transcripts themselves (i.e. utility of available formats). The paper based survey was attached to the end of each course's final examination, which was collected by an independent invigilator. The survey clearly indicated that participation was voluntary and would be treated anonymously. The professors were not present when the data was collected, nor did they have access to the results at any time.

Fig. 2 Student Survey

1. Were you provided information on the use of the Note Finder as a tool for accessing lecture notes online?	**Yes**	**No**
If "Yes":		
2. Did you find the Note Finder easy to use?	**Yes**	**No**
3. Did you find the Note Finder helpful?	**Yes**	**No**
4. When using Note Finder, did you use (circle one):	• **Keyword search** • **Phrase Table** • **Both**	
5. Would you recommend Note Finder to other students?	**Yes**	**No**

For the purposes of this initial analysis, surveys that were blank, incomplete, or contained a clear indication that Note Finder was not used, were not included. Out of 355 surveys collected, over seventy percent of the surveys were completed. Of those excluded, approximately 26% indicated the system was not used or were incomplete.

The results highlighted in Fig. 3 were extremely positive and consistent between both classes. As expected, the large majority of students found the tool intuitive and helpful. Most students reported using the keyword search or both Keyword and Phrase table. Only six percent reported using the phrase table in isolation. The most noteworthy class difference was the number of students that reported not using the system (Sociology 16% versus Anthropology 9%). This difference is believed to be attributed to the availability of an alternative access point (LAN) for the Sociology students.

Fig. 3 Note Finder Survey Results

	Anthropology		*Sociology*		*Total*	*Average*
Surveys Analyzed	141		111		252	
Aware - Note Finder	134	95%	107	96%	241	**95.63%**
Easy to Use	126	89%	98	88%	224	**88.89%**
Helpful	120	85%	95	86%	215	**85.32%**
Recommend	126	89%	99	89%	225	**89.29%**
Keyword	68	48%	59	53%	127	**50.40%**
Phrase	9	6%	6	5%	15	**5.95%**
Both	63	45%	46	41%	109	**43.25%**

A number of surveys contained confusing contradictions. For example, six respondents did not find the Note Finder easy to use or helpful, yet indicated they would recommend it to other students. Conversely, five students indicated the tool was easy to use and helpful, yet said that they would not recommend Note Finder to their peers.

Many respondents included written comments about their experience. Most were brief and reinforced either the individual's like or dislike of the system:

> *A source for those who find it hard to study from the textbook or their own notes….an easy way to find in-depth notes…very helpful if you need to find something quickly and it provides thorough information.*

Some anecdotal feedback spoke to difficulties with network access, availability of other resources, or redundancy. Although these are variables beyond our control, they contributed to some negative responses:

> *I tried to get onto NoteFinder but couldn't get in….it asked for my A number* (student ID) *and password but I couldn't figure out which one it wanted…I didn't have any reason for using Note Finder given the notes were available on the P:Drive*

Although the overall results were encouraging, the methodology used afforded exploratory insights at best. The unsophisticated survey provides a 'panoramic' view of user satisfaction, but does not enlighten developers on specific system features that require redesign or improvement. For example, although students reported high usability, there is no data available to target areas for improvement.

Another question that remains unanswered is whether it was the Note Finder system itself that generated the positive responses, the underlying SR generated transcripts, or the combination of both. Our initial hypothesis is that both are desirable in unison. When previously presented with SR generated transcripts without any advanced search tools, students still responded positively, but provided suggestions for enhancing usability. The introduction of intelligent search and retrieval tools should therefore improve the user experience.

6   Further Work

The Note Finder system discussed here provides an excellent framework for further improving the student experience. In particular, we hope to enhance notes to include hyperlinks between notes based on indexed phrases. Whenever students are reading the notes and come across an unfamiliar phrase, they will be able to follow the link and get more information from other class notes. The current format of the notes can also enable us to identify and tag (using new XML tags) parts of the document as *doclets* that describe a particular concept. These tags will make it possible for us to create a simple Semantic Web that can retrieve portions of the notes or "doclets" that are relevant to the user information needs, instead of retrieving the entire document.

Another interesting tangent is exploring social learning applications that can be facilitated through Note Finder. For example, students accessing SR transcripts could annotate sections they review, thus creating a searchable archive of peer transactions. A user could therefore see which sections were most frequently accessed by peers, perhaps indicating a socially generated measure of content importance.

7   Conclusion

Using SR to create multimedia transcripts available as "notes" for subsequent review seems to enhance learning opportunities in the university lecture environment. Coupled with intelligent search and retrieval tools like Note Finder, future generations of students could benefit from increased information accessibility. Determining whether the availability of SR generated multimedia transcripts and intuitive search tools improve academic performance is but one of many outstanding challenges facing researchers.

To tackle such questions, the Liberated Learning Consortium remains dedicated to continuing to investigate and advance SR applications that improve information accessibility. It is hoped that these efforts not only engender more sophisticated technologies, but also increase the necessary research collaborations that fuels such advancements.

References

1. Leitch, D., MacMillan, T. (2003). Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research. Saint Mary's University, Nova Scotia. Retrieved February 7, 2007, from http://www.liberatedlearning.com
2. Bain, K., Basson, S., Faisman, A., and Kanevsky, D, "Accessibility, transcription, and access everywhere". *IBM Systems Journal*. Volume 44, Number 3, 2005. http://www.research.ibm.com/journal/sj/443/bainaut.html
3. Salton, G., Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer , Addison-Wesley, 1989.

4. Salton, G. and BUCKLEY, C. 1988. "Term weighting approaches in automatic text retrieval", Information Processing Management , 24, 513-523.

5. Fagan, J.L., "The Effectiveness of a Nonsyntactic Approach to Automatic Phrase Indexing for Document Retrieval", Journal of the American Society for Information Science, 40(2), 1989, 115-132.

6. Mitra, M., Buckley, C., Singhal, A., and Cardie, C., "An Analysis of Statistical and Syntactic Phrases, Proceedings of RIAO97, Computer-Assisted Information Searching on the Internet , Montreal, Canada, 1997, pp. 200-214.

7. Anick, P.G., "The Paraphrase Search Assistant: Terminological Feedback for Iterative Information Seeking", Proceedings of SIGIR'99 , 1999. pp. 153-161.

8. Jones, S., "Phrasier: a System for Interactive Document Retrieval Using Keyphrases", Proceedings of SIGIR'99 , 1999, pp. 160-167.

9. Buckland M., "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies", D-Lib Magazine , http://www.dlib.org/dlib/january99/ buckland/01buckland.html, 1999.

10.Gutwin, C., Paynter, G.W., Witten, I.H., Nevill-Manning, C.G. and Frank, E., Improving browsing in digital libraries with keyphrase indexes , Technical Report, Department of Computer Science, University of Saskatchewan, Canada, 1998.

11. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C., and Nevill-Manning, C.G., KEA: Practical Automatic Keyphrase Extraction , Working Paper 00/5, Department of Computer Science, The University of Waikato, New Zealand, 2000.

12. Witten, I.H., "Browsing around a digital library", Proceedings of Australasian Computer Science Conference , Auckland, New Zealand, 1999, pp. 1-14.

13. Turney, P. D., "Coherent keyphrase extraction via Web mining", Proceedings of IJCAI'03 , 2003, 434-439.

14. Turney, P.D., "Learning algorithms for keyphrase extraction", Information Retrieval , 2(4), 2000, 303-336.

15. Desilets, A., de Brujin, B., and Martin. J. "Extracting keyphrases from spoken audio documents", SIGIR Workshop on Information Retrieval Techniques for Speech Applications , 2001, pp. 36-50.

16. Inkpen, D. and Désilets, A., "Extracting Semantically-Coherent Keyphrases from Speech" Journal of the Canadian Acoustics Association , 32(3), 2004, 130-131.

17. Bain, K. and Faisman, A. "Network ViaScribe Project". Liberated Learning Consortium. November 28, 2003.