

Yizkor Books: A Voice for the Silent Past

Jason J. Soo
Information Retrieval Lab
Illinois Institute of Technology
Chicago, Illinois
soo@ir.iit.edu

Rebecca J. Cathey
Advanced Information Tech.
BAE Systems
Arlington, Virginia
cathey@ir.iit.edu

Ophir Frieder
Information Retrieval Lab
Illinois Institute of Technology
Chicago, Illinois
frieder@ir.iit.edu

Michlean J. Amir
Archives Section
United States Holocaust
Memorial Museum
Washington, DC
mamir@ushmm.org

Gideon Frieder
School of Engineering and
Applied Science
George Washington University
Washington, DC
gfrieder@gwu.edu

ABSTRACT

Yizkor Book collections contain firsthand commemorative accounts of events from the era surrounding the rise and fall of Nazi Germany, including documents from before, during, and after the Holocaust. Prior to our effort, information regarding the content and location of each Yizkor Book volume was limited. We established a centralized index and metadata repository for the Yizkor Book collection and developed a detailed search interface accessible worldwide.

Categories and Subject Descriptors: H.4 [Information Systems Applications]: Miscellaneous

General Terms: Algorithms, Languages, Standardization

Keywords: Yizkor Books, Holocaust data archives, Historical documents

1. INTRODUCTION

Yizkor Books were published in thirteen different languages, across six continents, starting in the early 1940's and continuing to the present where the highest activity occurred during the 1960's and 1970's [1]. Yizkor Books provide firsthand accounts of the events in Europe that preceded, took place during, and followed the Second World War. These commemorative volumes contain documentation on destroyed communities and people who perished. Given their limited publication numbers, current fragile physical state, and to prevent destruction or theft, these volumes reside in isolation and under strict protection. One of the largest Yizkor Book collections is housed in the United States Holocaust Memorial Museum. Our system identifies, locates and describes digitally available Yizkor Books.

2. CHALLENGES

The diverse nature of and restricted access to the Yizkor Book collection complicates retrieval and introduces several unique challenges:

User Community: Readers of Yizkor Books vary in nature from knowledgeable, computer savvy research scholars to apprehensive individuals intimidated by computer

technology. Their interests range from finding specific family members to completing dissertations on specific topics. Some users are Holocaust survivors themselves.

Collection Language: Almost every book was written using multiple languages with the primary language of the book being 62% Hebrew, 24% Yiddish and the other 11 languages making up the remaining 14%.

Town Names: Some town names are popular; consequently, several towns in different areas of Eastern Europe have identical names. In some cases, town names and/or their spellings changed as many as 15 times.

Efficiency: In 1973, Yizkor Books already consisted of over 150,000 pages written by more than 10,000 authors [2]. Simultaneous efficient search access for Web users must be guaranteed; yet the host and back-up computers are cheap commodity workstations.

Content: The content of the Yizkor Books include narratives, detailed town by town listings of the history of the Jews in a town, their leaders, education, community self help organizations, maps of the town with the Jewish quarter detailed, lists of professionals, lists of victims, the story of the Holocaust in the town, as well as other aspects of life before, during, and after the Holocaust. Currently, our index and metadata cover nearly 700 books.

Resource Constraints: Computer resources are and always will be limited for this community. Currently, our system operates on a Dell Workstation (Intel Pentium 4 CPU 2.40GHz, 1GB RAM). All statistics were obtained from this environment.

3. SOLUTIONS

Using PHP and MySQL, we developed an AJAX based approach providing interactive assistance based on the data stored in the server. Our system provides:

Collection Languages: Users query the system for results in multiple languages without necessarily being literate in those languages. "On-the-fly" machine translators tend to produce inaccurate translations; thus, a complex set of parameters was designed to allow the user to select preferred categories to narrow their collection clusters. A user can also select the language they are looking for, a publication year, a publication town, author, etc., from the selection fields, thus retrieving their intended results.

Query	Misspelled	Results	Time(ms)
Virushov	No	1	3
Yizker	No	32	392
Sefer Vayslits:			
sefer 'edut			
ve-zikaron = Dos			
Vayslitser			
yizkor-bukh	No	1	168
Baron	Yes	max 20	433
Slavaka	Yes	8	35
Yidn in Plostik	Yes	1	55

Table 1: Query times

Town Names: Several towns have similar sounding names. Likewise, some towns have a diversity of spellings. Using predictive name expansion, via the AJAX techniques employed, users can select the correct town from the list of possible choices.

Efficiency: Efficiency was critical, even potentially at the expense of some feature expansion. As such, for the relational search component, many of the relations were denormalized. On average, query results are returned in less than 200ms.

Text Search: Currently, the metadata are plain text. As some searchable fields are correlated, a tagged organization supports a more accurate and efficient search. The integration of XML tagged document search with relevance ranking capability using relational technology is currently under development.

4. QUERY TIMES

To highlight the utility of the system, we measured the execution time for queries that range from single term, correctly spelled queries to multi-term, misspelled queries. The query terms are randomly selected book titles found within the database. Table 1 illustrates our results. These results demonstrate the efficiency of our system when the user entered exact query terms or was assisted by our suggestive search. It also shows misspellings are able to be identified and corrected for a minimal wait time.

5. FEATURES

The search interface embeds various features including but not limited to: automated query completion, ranked corrected term suggestions, and the restoration of deleted records. These features benefit both researchers and the administrative staff.

Restoration Feature: To simplify the operational management of the system, a restoration feature was developed to allow the staff to view deleted records and restore them.

Spell Checking: Regardless of the language proficiency of the user, many of the terms within the Yizkor Books, particularly geographical locations and personal names, are foreign to him/her. Thus, we crafted efficient simplistic rules to correct misspellings:

Algorithm	Phase 1	Phase 2	Phase 3
First	%aro%	%r%	-
Second	Ba%on	Ba%n	B%n
Third	%on	-	-
Fourth	Bar%	-	-
Fifth	B%n	-	-
Sixth	Ba%on	-	-

Table 2: Example Phases. Query: Baron

- 1 Replace first and last characters from query string with SQL search wild cards (%). Repeat string substitution until a sufficient confidence in the results returned is obtained or no additional results are found.
- 2 Replace middle character with SQL search wild card. Repeat string substitution.
- 3 Replace first half of query with SQL search wild card.
- 4 Replace second half of query with SQL search wild card.
- 5 Retain first and last characters, insert SQL search wild card in between retained characters.
- 6 Retain first and last two characters, insert SQL search wild card in between retained characters.

The query string is compared against each candidate generated using the rule set above with an upper bound on the number of potential candidates imposed. A scoring mechanism integrates the similarity scores derived in the comparisons. For each candidate, the derived scores generated by each of the rules are weighted and summed. A ranked list of candidates is presented to the user who can then use his/her judgment or base the decision on the rankings.

As with any database search, all correctly spelled terms are identified. Preliminary testing indicates, however, that even with our simplistic approach, for incorrectly spelled terms, the proper name is located 78% of the time as the top option and 9% of the time as the second option. Only in 18% of misspelled terms is the proper name not found.

For example, assume the desired query was "Baranovits" but the user inadvertently typed the title as "Baron." In such a case, our approach, after not retrieving any entries, would attempt to correct the spelling. Such correction using the above rules would derive the candidates illustrated in Table 2.

6. CONCLUSION

We deployed a search system to identify the location, content, and characteristics of digitally available Yizkor Books, a collection of significant historical and humanitarian interest. Prior to this effort, no centralized, extensive, in depth bibliography existed for this collection nor any other collection of its nature.

7. REFERENCES

- [1] M. Amir. *From Memorials to Invaluable Historical Documentation: Using Yizkor Books as Resource for Studying A Vanished World*. Annual Convention of the Association of Jewish Libraries, volume 36, 2001.
- [2] A. Wein. *Yad Vashem Studies on the Eastern European Catastrophe and Resistance*. Memorial Books as a Source for Research into the History of Jewish Communities in Europe, volume IX, 1973.