

Distributed Search for Large Collections of Symbolic Music

Alastair Porter

CIRMMT/McGill

28 July 2012

Outline

- 1 Introduction
- 2 Basic search
- 3 Speeding up search
 - Distributed search
- 4 Further work

Searching in music

The screenshot shows a web browser window with the URL `rodan.simsa.ca/8002/projects/17/dms`. The page title is "Searching CDN-Hsmu M2149.L4 (Salzines)". Below the title is a search bar with the text "edix" and a "Search" button. A message below the search bar reads "Result 1 of 6 for 'edix' (strict pitch sequence search)".

The main content is a zoomed-in view of a musical manuscript page. The page is titled "Lxxm" in red. It features four staves of music with square neumes on red four-line staves. The lyrics are written in black Gothic script below the staves. The first staff begins with a large blue initial 'G' and the text "lori a patri et fili o et spiri tui lan". The second staff begins with a large red initial 'G' and the text "loria patri et fili o et spiri tui sanct". The third staff begins with a large blue initial 'G' and the text "loria patri et fili o et spiri tui lan cto.". The fourth staff begins with a large red initial 'G' and the text "loria patri et fili o et spiri tui lan cto.". The search results are indicated by blue and black squares overlaid on the neumes.

At the bottom left of the manuscript view, there is a zoom level indicator showing "Zoom level: 4". At the bottom right, there is a page navigation bar showing "Page 1 of 1" and a "CD" icon.

Searching in music

The screenshot shows a web browser window with the URL `ddm.music.mcgill.ca/liber?type=names&query=edcdeee&p=117&z=3`. The page title is "Search the Liber usualis". Below the title, a search bar contains the text "edcdeee" and a "Search" button. To the right of the search bar are "Clear" and "next" buttons. Below the search bar, it says "Result 1 of 40 for edcdeee". To the left of this text is a "previous" button and to the right is a "next" button. Below these buttons is a "Go to page" field with a "Go" button. Below the search results, there is a musical score snippet. The text of the score is: "in excél-sis. Benedictus qui vé-nit in nómine Dómi-ni. Hosánna in excél-sis. XII. e. 4. **A** -gnus Dé-i, * qui tóllis peccáta mún-di : mi-se-ré-re nó-bis. Agnus Dé-i, * qui tóllis peccáta mún-di : mi-se-ré-re nó-bis. Agnus Dé-i, * qui tóllis peccáta mún-di :". The musical notation is on a five-line staff with a treble clef and a key signature of one flat. The score is displayed on a light-colored background with a dark border.

Hosted by the Distributed Digital Music Archives & Libraries Lab of McGill University. Contact us with questions or comments

Current infrastructure

- Fast search for musical phrases
- Highly accurate transcriptions
- Search over monophonic pitches
 - Plus contour, neume names
 - (no rhythm, polyphony)
- Highlight results over top of original image source

Equivalents in text searching

Beethoven

temperaments—the father being as sour in disposition as the mother was sweet. One result of this was that young Ludwig adored his mother and feared his father—the latter a rather unfortunate circumstance, inasmuch as he was to come a good deal under the paternal influence.

The Beethovens were very respectable but poor. The “Van” in the name was no sign of nobility, and less than $\frac{1}{2}$ so per annum—the father’s salary in the Elector of Cologne’s Chapel—was the slender income upon which the family had to subsist. Little wonder that the head of the household ruled with a stern—even cruel hand. Who could be complainant in face of such a condition of ways and means? That he fell betimes into irregular habits, though eminently indefensible, was not surprising. Granted, then, he was a severe cross-grained parent, with an insatiable temper, faults which made the humble house less comfortable than it should have been for the great mind sharing its roof. Never would it be Beethoven’s portion, in after years, to enjoy that priceless blessing, the coming retrospect of a happy childhood.

The burden of poverty which the Beethovens supported was not altogether an unmitigated evil, inasmuch as but for stress of domestic conditions this second child would *Earlier not, probably, have received so large a share of Training the parent’s musical ministrations. Inspired by what had been accomplished by the father of Wolfgang Amadeus Mozart as a musical “wonder,” however, Beethoven senior set about the shaping of another harmonic prodigy. The incentives were eminently favourable. Poverty stared in the family’s face, and here was a child with pronounced musical ability. His com-*

A Promising Pupil

mercial value was gauged therefore; and to this circumstance, mainly, the world owes its heritage of Beethoven. The father took the little fellow in hand, and kept him to study practice and exercises almost unceasingly. Some small accompaniment of general education crept in through one of the common schools, but this terminated ere the boy had reached his thirteenth year.

Beethoven was four years old when he began to study music, and at the age of nine he had learned all that his father had to teach him. This consisted of instruction in the piano, violin and harmony, all which the quiet, gave child learned readily—though not without frequent reproach and cuff from the harsh parent. He came also under the notice of Pfeiffer and Zambona. The former, a boon companion of the father’s, continued the boy’s musical education—Zambona, meanwhile, teaching him some Latin, French and Italian. Subsequently, Van den Eeden and Neefe took young Beethoven in charge. They instructed him in organ-playing and musical theory; and from

Search: The First Biography

books.google.ca/books?id=6u6dZaEg2w0C&pg=PA10&dq=Beethoven+was+four+years+old+when+he+began+to+study+music+and+at+the+age+of+nine+he+had+learned+all+that+his+father+had+to+teach+him+this+consisted+of+instruction+in+the+piano+violin+and+harmony+all+which+the+quiet+gave+child+learned+readily+though+not+without+frequent+reproach+and+cuff+from+the+harsh+parent+He+came+also+under+the+notice+of+Pfeiffer+and+Zambona.+The+former,+a+boon+companion+of+the+father’s,+continued+the+boy’s+musical+education+Zambona,+meanwhile,+teaching+him+some+Latin,+French+and+Italian.+Subsequently,+Van+den+Eeden+and+Neefe+took+young+Beethoven+in+charge.+They+instructed+him+in+organ+playing+and+musical+theory+and+from

books.google.ca/books?id=6u6dZaEg2w0C&pg=PA10&dq=Beethoven+was+four+years+old+when+he+began+to+study+music+and+at+the+age+of+nine+he+had+learned+all+that+his+father+had+to+teach+him+this+consisted+of+instruction+in+the+piano+violin+and+harmony+all+which+the+quiet+gave+child+learned+readily+though+not+without+frequent+reproach+and+cuff+from+the+harsh+parent+He+came+also+under+the+notice+of+Pfeiffer+and+Zambona.+The+former,+a+boon+companion+of+the+father’s,+continued+the+boy’s+musical+education+Zambona,+meanwhile,+teaching+him+some+Latin,+French+and+Italian.+Subsequently,+Van+den+Eeden+and+Neefe+took+young+Beethoven+in+charge.+They+instructed+him+in+organ+playing+and+musical+theory+and+from

Google

Search: Mozart

Books

Find all the books to read

GET MOST BOOKS

No eBook available

Reprints Books

Amazon.ca

Check out this book

Find in a library

All sales +

Mozart: The First Biography

By Tracy Storr

Beethoven

Buy

About this book

LIFE OF MOZART

AS THE SWEET enchantment of Mozart’s harmony and the overwhelming stream of his feeling never fail to raise his listeners to ecstatic heights, so the student of human nature cannot but admire the early development and rapid maturity of his creative genius.

From both standpoints *Mozart* inspires so great an interest that I beg leave to hope that this biographical sketch will prove a not unwelcome offer-

Some terminology

- Corpus
 - A collection of documents to search over
- Document
 - A single source of content to look at (page of music, musical work, book)
- Query
 - What we are searching for

Linear search

Look at each document in the corpus and see if part of the document matches the query

Problems:

- As you get a larger corpus this takes longer and longer to complete
- Naive search has no optimisations or caching—if you perform the same query twice in a row, the search is performed twice (wasteful)

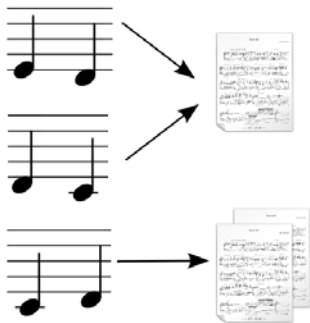
Speeding up search

- Process each document to extract musical information before making it available to search
- Split a document into sequential tokens of length n



Speeding up search

- Store as an *inverted index*
- Like a book index, look up all pages where a sequence of notes appears
- Can search thousands of pages of music in 10s of milliseconds, millions of pages of text in seconds (compared to hours)



Limitations of indexes in music search

- Text is relatively easy to index
- Music has many dimensions that need to be indexed
 - Pitch
 - Rhythm
 - Vertical intervals (Polyphony)
 - High-level relationships
- An n-gram search approach for music can work well for monophonic sequences of notes
- Difficult to make exploratory searches without doing a naive linear search

Searching with musical meaning

Our goal: To be able to perform musically complex queries on large corpora of music in a timely manner

- With indexed search the index needs to be computed ahead of time
- You can't predict all the types of search that might be made
- For example, search for instances of a cadence over a four-part score

Distributed search

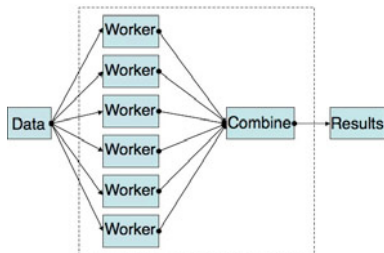
- Revisit linear search
- Use musically aware search software (music21/humdrum)
- Can we make it faster?
 - Get lots of computers to perform the search in parallel



<http://hadoop.apache.org/>

Hadoop

- Framework for parallel computing
- Split up the corpus evenly over a large number of computers
- Each computer processes a “mini corpus”
 - Could search a for a simple sequence of notes
 - or perform a more complex humdrum/music21 query
- Combine results once all computers have finished



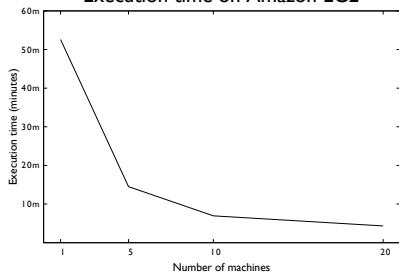
Our experiment

- Use two clusters
 - Amazon EC2
 - CLUMEQ (Consortium Laval, Université du Québec, McGill and Eastern Quebec)
- Input: 234,000 pages of music encoded in MEI format
- Distribute over 1, 5, 10, 20 computers
- Expect that using 10 machines takes $\frac{1}{10}$ th the time of 1
- Still using the same number of *CPU hours*



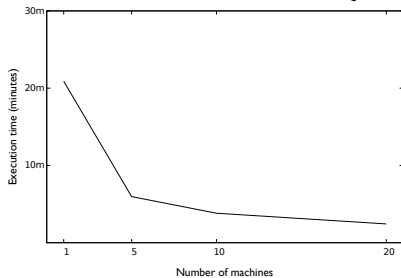
Results

Execution time on Amazon EC2



Machines	Time	Speedup
1	3156s	-
5	871s	3.56×
10	417s	7.43×
20	261s	11.86×

Execution time on CLUMEQ



Machines	Time	Speedup
1	1254s	-
5	358s	3.50×
10	229s	4.49×
20	146s	8.58×

What's next?

- We can create an index for common queries
 - e.g., key
- Perform an indexed query to limit the number of documents to search then perform an in-depth search
 - Get all works in the key of C
 - Analyse structure

Thanks



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

