

## INDEXING

---

- Time: 40.3 hrs (5-6 min per 100 MB on avg )
- Size: 4.6 GB

## SECONDARY INDEX

---

- Procedure: N is width where N total size of vocabulary
- Time: 2 min 31.862 seconds ~ 2 min approx.
- Size: 48KB

## TITLE EXTRACTION-INDEXING TITLES

---

- Index Size: 377MB
- Time: 19min 55 sec ~ 20min approx

## TITLE SECONDARY INDEX

---

- Size: 144KB
- Time: 1m9.033 seconds ~ 1 min approx

## RANKING

---

- Input-1: T : Thierry T : henry
  - Time: 372 ms
  - Results:
  - *13035847 Wikipedia:Peer review/Thierry Henry/archive2*
  - *220160 Thierry Henry*
  - *25316130 Wikipedia:Featured article review/Thierry Henry/archive1*
  - *9334631 Thierry henry*
  - *28148845 Thierry Henri*
  - *25255794 Thierry Henry handball incident*
  - *5194828 Henry Thierry*
  - *5963046 Wikipedia:Featured article candidates/Thierry Henry/archive1*
  - *11966736 Thierry Daniel Henry*
  - *12377281 Wikipedia:Peer review/Thierry Henry/archive1*
  - *17074470 Thierry D. Henry*
  - *22221117 File:Thierry Henry 2008 small.jpg*
  - *11764107 Henry Henry*
  - *26042338 St. Henry's Catholic Church (St. Henry, Ohio)*
  - *23427075 Louis-Henri Mercier (Henri Goegg)*
- Input-2: T : Sachin T : Tendulkar
  - Time: 279 ms
  - Results:
  - *15866602 Wikipedia:Featured list candidates/List of Test cricket centuries by Sachin Tendulkar*
  - *24055426 Template:User Sachin Tendulkar*
  - *21221965 List of ODI Awards for Sachin Tendulkar*
  - *16052267 List of International cricket centuries by Sachin Tendulkar*
  - *15822979 Wikipedia:Peer review/List of ODI Awards for Sachin Tendulkar*
  - *35877360 Sachin Tendulkar records*

IRE REPORT MINI PROJECT -1  
200902005

- 15622033 *Wikipedia:Featured list candidates/List of international cricket centuries by Sachin Tendulkar*
  - 9949732 *Sachin r tendulkar*
  - 20408026 *Wikipedia:Featured article candidates/Sachin Tendulkar/archive1*
  - 10214916 *List of ODI awards for Sachin Tendulkar*
  - 23318954 *Achievements of sachin tendulkar*
  - 10214931 *List of ODI Man of the Match Awards for Sachin Tendulkar*
  - 11205968 *Wikipedia:Peer review/List of ODI Awards for Sachin Tendulkar/archive1*
  - 10214926 *List of ODI Man of Match Awards for Sachin Tendulkar*
  - 35099061 *File:Sachin Tendulkar cropped.jpg*
  - Input-3: I : greek B : python C : mythology
    - Time: 464 ms
    - Results:
    - 594 *Apollo*
    - 19272020 *Gaia (mythology)*
    - 18574 *Leto*
    - 20634 *Minotaur*
    - 17528421 *Bashe*
    - 2561894 *Melissa*
    - 99498 *Anansi*
    - 7951 *Delphi*
    - 26795856 *Wikipedia:Bot requests/Archive 35*
    - 491152 *Merman*
    - 30256564 *National Geographic Channel Greece*
    - 83148 *Danaë*
    - 29067 *Satyr*
    - 6697 *Cerberus*
    - 7401 *Centaur*
  - Input-4: T : ice B : glacial B : jura B : drumlins
    - Time: 461 ms
    - Results:
    - 15361 *Ice age*
    - 44298 *Drumlin*
    - 12463 *Glacier*
    - 557672 *Last glacial period*
    - 27462 *Geography of Switzerland*
    - 182061 *Geography of the Interior United States*
    - 2308763 *Swiss Plateau*
    - 1407789 *Ain (river)*
    - 461435 *Drumlin field*
    - 17288465 *Ice-sheet dynamics*
    - 31594707 *John Mercer (archaeologist)*
    - 20994108 *List of nature parks in Germany*
    - 21245728 *Geology of Saskatchewan*
    - 36695569 *Vooremaa*
    - 706216 *Jökulhlaup*
-

## FILE USAGE

---

Searchindex40gb	Index file
Secondaryindex40gb	Secondary index file
Titles	Tiles file
Titlesec	Secondary index for titles
Search.py	Indexing part
Rank.py	Ranking part
Buildsecondaryindex.py	Building secondary index for “searchindex40gb” file
Titlesecondaryindex.py	Building secondary index for “titles” file
Getdoccount.py	Getting document count
Base.py	Base 62 to decimal conversion
Finalstoplist	Stop word list for whole process