

Sommario

Repository.....	1
Dati corpus	1
Analyzer utilizzati	1
Tempi di indicizzazione	2
Query effettuate.....	3

Repository

Repository: <https://github.com/paolo-di-simone/homework2>

Dati corpus

Numero di file utilizzati per l'indicizzazione: **8600**

Numero medio di parole per file: **186.94814**

Corpus: Ohsumed collection (available at <ftp://medir.ohsu.edu/pub/ohsumed>): it includes medical abstracts from the MeSH (Medical Subject Headings, is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed) categories of the year 1991.

Tipologia file: Pathological Conditions, Signs and Symptoms

Analyzer utilizzati

Sono stati utilizzati 2 Analyzer distinti:

- Per i nomi dei file è stato utilizzato uno **StopAnalyzer** con **stop word "txt"**: questo analyzer permette di
- Per i contenuti dei file è stato utilizzato un **EnglishAnalyzer**.

Lo StopAnalyzer suddivide il testo in token eliminando qualunque carattere non sia una lettera (come numeri, spazi, trattini molto utilizzati per assegnare il nome ad un file), inoltre porta tutti i caratteri in minuscolo.

Oltre a questo, accetta una lista di stop word da scartare; come unica stop word è stata utilizzata "txt", cioè il formato del file.

Uno StandardAnalyzer non sarebbe andato bene per i nomi dei file, dal momento che nomi tipo "parola1-parola2" sarebbero stati considerati un solo token, mentre con uno StopAnalyzer vengono suddivisi in due token. Per il contenuto dei file è stato utilizzato un EnglishAnalyzer, dal momento che i documenti elaborati sono tutti scritti in inglese.

Tempi di indicizzazione

Indicizzazione 1 effettuata in: 7.972(s), **7972(ns)**

Indicizzazione 2 effettuata in: 5.602(s), **5602(ns)**

Indicizzazione 3 effettuata in: 5.131(s), **5131(ns)**

Indicizzazione 4 effettuata in: 5.104(s), **5104(ns)**

Indicizzazione 5 effettuata in: 5.067(s), **5067(ns)**

Indicizzazione 6 effettuata in: 4.983(s), **4983(ns)**

Indicizzazione 7 effettuata in: 5.048(s), **5048(ns)**

Indicizzazione 8 effettuata in: 5.129(s), **5129(ns)**

Indicizzazione 9 effettuata in: 5.256(s), **5256(ns)**

Indicizzazione 10 effettuata in: 5.033(s), **5033(ns)**

Tempo medio indicizzazione (10 indicizzazioni effettuate): **5432.5(ns)**

Query effettuate

Query 1: **minimize the accumulation of calcium in damaged cells**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: minimize the accumulation of calcium in damaged cells

Documento 3528: IN.txt (9.832636)

Documento 2054: DISEASE well more-Case influenzae-Surveillance Serotypes Infection-Most Haemophilus.txt (5.799809)

Documento 6474: Several94.txt (5.686283)

Documento 6402: SEPTIC_State-a New-and_Be.txt (5.2486176)

Documento 1777: cusp-Bacterial-meningitis-that Cause Of Jersey_That streptococcal population.txt (5.1521416)

Documento 2113: due case That 5 Multistate Her endocarditis_hand TREATMENT meningitis.txt (5.1310453)

Documento 3647: Influenzae VARIED96 in46 antibiotic Neisseria_was-she Invasive-meningitis11 than.txt (5.1270022)

Documento 1576: causes was optimal in-OPTIMAL_IN12 United.txt (4.717538)

Documento 3759: in_in Of2.txt (4.693012)

Documento 1838: Data_B SUGGESTING.txt (4.616284)

Tempo di risposta query: 2.032 (s), 2032 (ns)

Query 2: **laryngotracheoplasty**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: laryngotracheoplasty

Documento 319: ACCURATE76-for-YEARS-Medical-streptococcus of THAT DISEASE-to.txt (6.0993414)

Documento 3777: in_Was Meningitis Half endocarditis.txt (5.0842686)

Tempo di risposta query: 1.214 (s), 1214 (ns)

Query 3: **athletes**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: athletes

Documento 1835: data24-the-NEONATAL.txt (5.242964)

Documento 7945: tolerance MOST Serotypes11 MENINGITIS Septic37 several Despite-At.txt (4.985754)

Documento 2482: fashioned-the_be26-WAS IMPROVEMENTS.txt (4.8967924)

Documento 1404: Calf Data second DATA common.txt (4.869658)

Documento 1766: cusp SURGERY rates.txt (4.827862)

Documento 8184: varied More-than-state Of Report-B.txt (4.735405)

Documento 4121: meningitis several-people accurate reported group.txt (4.423935)

Documento 1315: by group Common SECOND8 She40_greater_EMERGENCY-VARIED.txt (4.41967)

Documento 4795: new-Meningitis_Resultant_in SURVEILLANCE-by.txt (4.4125805)

Documento 1161: Bacterial group with_in-study-Organisms operation-and_WAS.txt (4.3530827)

Tempo di risposta query: 1.189 (s), 1189 (ns)

Query 4: **symptoms of paralysis**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: symptoms of paralysis

Documento 2104: disease_PROJECT.txt (5.348133)

Documento 2550: followed 1986.txt (5.1970167)

Documento 6962: suggesting11 COMMON also-AND CAUSED-Nine_influenzae75 second influenzae71 Later.txt (4.567223)

Documento 5535: other of of97 Washington United mortality.txt (4.4975824)

Documento 158: A B_common THE that6 A.txt (4.2153964)

Documento 3604: INFECTION_State Optimal State State in.txt (4.0849733)

Documento 1954: Developed project STREPTOCOCCUS_Infection INFECTION Of-be invasive CONCURRENT.txt (4.0340366)

Documento 5685: Pericardium the IIII60_REQUIRED surgical63 in Since-34.txt (3.982093)

Documento 3892: LOWER Be studies exercise47-later state-neonatal Optimal.txt (3.938995)

Documento 5421: Operation the_to A.txt (3.938995)

Tempo di risposta query: 1.161 (s), 1161 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: "symptoms of paralysis"

Documento 3604: INFECTION_State Optimal State State in.txt (4.0849733)

Tempo di risposta query: 1.121 (s), 1121 (ns)

Query 5: **optic nerve**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: optic nerve

Documento 5504: organisms14 Of_of-cusp.txt (7.4997606)

Documento 1587: causes32.txt (7.3911605)

Documento 5395: ON years-followed Most Aortic performed surgery IN-years.txt (7.2165337)

Documento 3784: Jersey MENINGITIDIS SINCE.txt (7.126174)

Documento 7529: the Nine of.txt (6.8956895)

Documento 2702: from Washington.txt (6.7329006)

Documento 475: an DEVELOPED STUDY LATER-group MENINGITIS-rates THAN_and63.txt (6.6712112)

Documento 4793: NEW-B-MEDICAL19 meningitidis19.txt (6.1090226)

Documento 5533: Other NEONATAL-Study Also97 Caused.txt (6.1090226)

Documento 1292: be-Of_due-Developed-cause From61.txt (5.8134604)

Tempo di risposta query: 1.11 (s), 1110 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: "optic nerve"

Documento 5504: organisms14 Of_of-cusp.txt (7.49976)

Documento 5395: ON years-followed Most Aortic performed surgery IN-years.txt (7.2165337)

Documento 3784: Jersey MENINGITIDIS SINCE.txt (7.126174)

Documento 1587: causes32.txt (6.7419753)

Documento 475: an DEVELOPED STUDY LATER-group MENINGITIS-rates THAN_and63.txt (6.6712112)

Documento 7529: the Nine of.txt (6.582134)

Documento 2702: from Washington.txt (6.4940104)

Documento 4793: NEW-B-MEDICAL19 meningitidis19.txt (6.109022)

Documento 5533: Other NEONATAL-Study Also97 Caused.txt (6.109022)

Documento 240: A-antibiotic PNEUMONIAE-varied ARTHRITIS79.txt (5.5569706)

Tempo di risposta query: 1.115 (s), 1115 (ns)

Query 6: **microneurosurgery**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: microneurosurgery

Documento 7203: than Meningitis_A Her-DETECTION76 those CONCURRENT Multistate-than_The.txt (4.2784286)

Documento 5579: Overall Washington The-Followed.txt (3.9379792)

Tempo di risposta query: 1.283 (s), 1283 (ns)

Query 7: **main causes of bilateral blindness**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: main causes of bilateral blindness

Documento 5528: other IN51-Caused_More TO That early_the_AND.txt (9.42498)

Documento 7952: TO_Meningitis-was.txt (5.3818817)

Documento 7947: tolerance_Regurgitation OF31 surgical.txt (5.1308517)

Documento 1484: cause OBTAINED the_neonatal-multistate Causes in Meningitis Group.txt (4.8857017)

Documento 1953: DEVELOPED Preparation Million_STUDIES had FROM-tolerance.txt (4.5157013)

Documento 3818: laboratory-based IIII.txt (4.4822617)

Documento 7675: the-than_cusp_hand.txt (4.4665403)

Documento 1324: by INVASIVE project antibiotic_Disease-Preserved regurgitation_Mortality.txt (4.412652)

Documento 1874: DESPITE55-People vaccine91 Varied HALF.txt (4.188304)

Documento 7546: The POPULATION a_that.txt (4.0364304)

Tempo di risposta query: 1.125 (s), 1125 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: "main causes of bilateral blindness"

Documento 5528: other IN51-Caused_More TO That early_the_AND.txt (6.3529716)

Tempo di risposta query: 1.128 (s), 1128 (ns)

Query 8: **clinical parameters of the status of general health**

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: clinical parameters of the status of general health

Documento 7585: the Time streptococcal82 a25 Antibiotic-That.txt (7.455798)

Documento 6966: SUGGESTING86 in A-AN her.txt (7.401168)

Documento 8395: Washington-of Improvements-bacterial emergency Her For to.txt (6.5062456)

Documento 5083: of Of.txt (6.3643093)

Documento 6792: streptococcus_Treatment21_more-CAUSES TIME disease disease NEONATAL caused And.txt (5.938784)

Documento 2941: GROUP85 COMMON8 THAT-B MENINGITIS meningitidis treatment.txt (5.812868)

Documento 8331: was-must Streptococcus WELL Reported0 in34 of7-IN cusp73 data74.txt (5.7048864)

Documento 2372: endocarditis_NEW-Streptococcal exercise-case.txt (5.6886916)

Documento 598: and IMPROVEMENTS.txt (5.5326805)

Documento 2014: Disease during0 persons organisms regurgitation surgical-in-accurate_AORTIC-common.txt (5.342047)

Tempo di risposta query: 1.141 (s), 1141 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 2

Cerca: "clinical parameters of the status of general health"

Documento 7585: the Time streptococcal82 a25 Antibiotic-That.txt (7.455798)

Tempo di risposta query: 1.123 (s), 1123 (ns)

Query 9: **heart disease**

Campo su cui fare query, nome file (1), contenuto (2): 1

Cerca: heart disease

Documento 2030: disease new-Of-A-FOLLOWED16.txt (2.6703746)

Documento 2039: disease several.txt (2.6703746)

Documento 2316: endocarditis IN surgical_Multistate disease disease Of41 SIGNIFICANTLY39-from.txt (2.4124846)

Documento 6792: streptococcus_Treatment21_more-CAUSES TIME disease disease NEONATAL caused And.txt (2.4124846)

Documento 999: At_3430_Tolerance disease B.txt (2.3664515)

Documento 2008: disease And That.txt (2.3664515)

Documento 2035: disease Optimal PRESERVED68-Other-Than27.txt (2.3664515)

Documento 2047: disease than NEISSERIA19.txt (2.3664515)

Documento 2048: disease The group_Infection.txt (2.3664515)

Documento 282: a13 a disease Invasive-Jersey.txt (2.1246397)

Tempo di risposta query: 1.214 (s), 1214 (ns)

Query 10: rates disease bacterial

Campo su cui fare query, nome file (1), contenuto (2): 1

Cerca: rates disease bacterial

Documento 5985: Rates disease Bacterial-OTHER-The SUGGESTING.txt (3.6908503)

Documento 5832: preparation DISEASE rates Must Disease-population Influenzae_Bacterial-Streptococcus_people.txt (3.3436522)

Documento 2057: Disease-and RATES.txt (3.3003774)

Documento 6682: Streptococcus Bacterial Rates-also neonate disease PROJECT Washington.txt (3.238947)

Documento 6018: Rates-In73 aortic disease86.txt (3.032809)

Documento 130: 5-disease bacterial Than8.txt (2.904974)

Documento 5747: pneumoniae BACTERIAL-SEPTIC MUST Rates bacterial.txt (2.8879485)

Documento 3930: MAY BACTERIAL_DISEASE COMMON94 SEVERAL Rates-PERICARDIUM-exercise suggesting and.txt (2.885634)

Documento 1260: bacterial_RATES_streptococcus influenzae.txt (2.8763285)

Documento 6036: rates_people Rates30-and Washington_organisms STUDY26-disease.txt (2.8237772)

Tempo di risposta query: 1.062 (s), 1062 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 1

Cerca: "rates disease bacterial"

Documento 5985: Rates disease Bacterial-OTHER-The SUGGESTING.txt (3.6908503)

Tempo di risposta query: 0.641 (s), 641 (ns)

Query 11: **project study**

Campo su cui fare query, nome file (1), contenuto (2): 1

Cerca: project study

Documento 3601: Infection63 Bacterial79 Most Study during project.txt (3.0713482)

Documento 5757: pneumoniae project STUDY developed33 PEOPLE The Concurrent.txt (2.8710606)

Documento 4700: NEONATAL69 PROJECT And Study new_That-Occurred New.txt (2.6952958)

Documento 5937: project.txt (2.626495)

Documento 4993: OF during Study-in61 Haemophilus ANTIBIOTIC ON-project must.txt (2.5398107)

Documento 1179: Bacterial OF early76 States99_endocarditis AND52-neonatal Project cusp_Study63.txt (2.4012856)

Documento 870: AORTIC project84.txt (2.3723516)

Documento 903: Aortic-PROJECT.txt (2.3723516)

Documento 1259: BACTERIAL_PROJECT.txt (2.3723516)

Documento 2104: disease_PROJECT.txt (2.3723516)

Tempo di risposta query: 0.568 (s), 568 (ns)

Campo su cui fare query, nome file (1), contenuto (2): 1

Cerca: "project study"

Documento 5757: pneumoniae project STUDY developed33 PEOPLE The Concurrent.txt (2.8710608)

Tempo di risposta query: 0.562 (s), 562 (ns)