

Group: Cara Failer, Siddharthen Sridhar  
Class: CS121/INF141  
Professor: Crista Lopes  
Assignment: A3 Milestone 2

# Assignment 3

## Milestone 2

# Introduction

With the creation of the inverted index from Milestone 1, our team has implemented the search retrieval function of the search engine. This added feature reads all files provided in the 'DEV' folder, and discovers whether specific terms can be found. This report documents the specific terms discovered in the documents and the frequency of these terms. In particular, our team is documenting the top five documents the terms were found in, and their frequency. A screenshot has also been provided of the search retrieval process

# Final Report

## Top URLs for Each Query

### Top 5 URLs for 'Cristina Lopes'

URL	Frequency
chenli_ics_uci_edu\0dd51496522b853a3cd23b185ecdd452cb906d3ec82819f2a4c3bce2d39cefe4.json	1
flamingo_ics_uci_edu\ae2a95d3d6d22c4be16834e4415790ac945fdd42f9dcaa603bd3d8fe68d89494.json	1
mondego_ics_uci_edu\32234f92a79c006e6027f831568fc099f6ed384a07d4732596acb45fd966dd7d.json	1
mondego_ics_uci_edu\6c647534f5a28e31cb903c544fc38a8b1cc28b3ebd0ef08dd81ba35ef401732a.json	1
mondego_ics_uci_edu\9f07aa7921ba582578986346e4f2a9e10f12becd4eb09ae11c2b9ecda1ed0a3f.json	1

### Top 5 URLs for 'Machine Learning'

URL	Frequency
archive_ics_uci_edu\013dca56ce14889be0e029a7a0673ddb1f7c945e08529403b73f347798cda7.json	1
aiclub_ics_uci_edu\8ef6d99d9f9264fc84514cdd2e680d35843785310331e1db4bbd06dd2b8eda9b.json	1
archive_ics_uci_edu\00db3b44bb015e65e2675829c71f4269b5d9fbaf282e89fb931f492b35dca268.json	1
archive_ics_uci_edu\00f8383295fe6bd7304d4728beb798920dccb576f2832ea3549904384ac3655d.json	1
archive_ics_uci_edu\0191a0170c9a3b6969462ddd8d7ff512ff516e69d4432e0fdb75766176515cb.json	1

## Top 5 URLs for 'ACM'

URL	Frequency
www_ics_uci_edu\d3338037b6de42a9591a911b6ece00d09f342269867a82237a66737e9974875b.json	139
www_ics_uci_edu\aff238d8493a383c284f875e75ac4566c1fd1f6e594438d25719901ac3e5fe2a.json	124
www_ics_uci_edu\59988e844aed0070085a78cad5b4f39410cf81c42b673c9a30c90d5297938d0e.json	89
www_ics_uci_edu\e04e852700abfa9448747f3eb7bb59236023df33989dcc21fcf5746c11fac3ee.json	81
www_ics_uci_edu\8608c79e77054cb0e4853301f1a03f1a92634b678d32ac0a6ab7b0d3cd46e1f6.json	75

## Top 5 URLs for 'Master of Software Engineering'

URL	Frequency
mswe_ics_uci_edu\007f1dcab3731c49a22f8ff0d1a24ec0c26759650b01b2d5dbe3a787865e13ea.json	1
mswe_ics_uci_edu\067c7f73631864772190487734b60867cbc9340c93b6d1821de7306d25826029.json	1
mswe_ics_uci_edu\1c1131431fd2a251600dd0548975ff7c55dc5549ef805553ceb4e0f90ee19350.json	1
mswe_ics_uci_edu\1efc6180dc3995d65127a94a74265604aeb1d21e8e803ba86f35be473004e2c2.json	1
mswe_ics_uci_edu\2a682642ab15d837cadceafb06426b1729bcfe1beb7f61e040a79c0b49ec3167.json	1

# Screenshot

Below is a screenshot of our current interface actively finding the frequency of specific terms. In this case, it is finding the frequency of the term, “ACM”.

```
Document ID: wics_ics_uci_edu\T0000//T00//a54/a/y2e1e0c0e70/13//T00c0e001T00103T013C1YTC54e3.json (Frequency: 1)
Document ID: wics_ics_uci_edu\fc1cfdde55825d20da70fadf003cc13694c5b0d8c7e32a89a2f38fcf18ffa515.json (Frequency: 1)
Document ID: wics_ics_uci_edu\fc7c29442224355cbdc2dd4d406872f6f58fe888d9ff8506cc07b5a0bf0deeb4.json (Frequency: 1)
Document ID: wics_ics_uci_edu\fe38125b91681c475ff0557292934d36a27f5dc250f4ec942e2e761d36995b4d.json (Frequency: 1)
Document ID: wics_ics_uci_edu\fe558891e6de933e21e92fbca7b351e0364c0354a12f2c7a5e0a2213440daa10.json (Frequency: 1)
Document ID: wics_ics_uci_edu\ff79394eba80c9bb5528522b758b663e64c533f23da35793b29172e3efba287a.json (Frequency: 1)
Document ID: wics_ics_uci_edu\ffd3e2dc476c982dc74925c3b13a7a0ff950308cbd3d27b33f362cb05717fd8e.json (Frequency: 1)
Document ID: www-db_ics_uci_edu\0ce97c31a62b7b8bb7205b10952be656f76f759b4517500bfe035fb10e0a4fe8.json (Frequency: 3)
Document ID: www-db_ics_uci_edu\16c4d46a219e4961a76bb1e1bc7b5cd2d812c3ec2580baf91f0a4ad89cc0d208.json (Frequency: 2)
Document ID: wics_ics_uci_edu\fffa3f1851b9ed63e35573eb3a42c0f5b71011588545bfff3c72f635d41c6909a.json (Frequency: 1)
Document ID: www-db_ics_uci_edu\2b8cbd2476f476856e9b95539c6b9b9f3e34841863c922f15dce4bbebb3ed245.json (Frequency: 5)
Document ID: www-db_ics_uci_edu\45f1d9727a46c4546b8d9fd965c03840f2842b698e9431dfd1d4ac9b5dfe35b2d.json (Frequency: 2)
Document ID: www-db_ics_uci_edu\63964bdda89cbd23703b9de565176200c86c0df86b2e84875ec075c222190c08.json (Frequency: 3)
Document ID: www-db_ics_uci_edu\6df49a408b214529489f0d474b26409d484b1a15a94e4e72203246885fadfb28.json (Frequency: 2)
Document ID: www-db_ics_uci_edu\c922831c5c20684fa38935a86ea2bea6df6bebeefcfbf63cf440d27ac7e0948f.json (Frequency: 2)
Document ID: wics_ics_uci_edu\ffff27ecd9cee22609edb3f93f24cd2cf69d99b82e701e197f55c2f0352f3039.json (Frequency: 1)
Document ID: www-db_ics_uci_edu\f705bbf4912eae950a290d59e48f7570272557b5e214d2493acc35329f10d59a.json (Frequency: 5)
Document ID: www-db_ics_uci_edu\e93aeb75f266e72b610a63436531d3d9e5a3e78ea161c3a6c233eb0e01beba9a.json (Frequency: 5)
Document ID: www-db_ics_uci_edu\fc24f5b082502a614240e47f1461ee43bfd1820e49452f11c74b6a31b4a28eb.json (Frequency: 5)
Document ID: www_cs_uci_edu\0107910665eb15542a4d026abf5e02a0b563c59335c0a7fdf385f33b39e59ead.json (Frequency: 2)
```