UNIVERSITY OF CYPRUS
Computer Science Department

ΕΠΛ 660 – Information Retrieval and Search Engines

Lab8: Elasticsearch

Lab instructor: Pavlos Antoniou

http://www.cs.ucy.ac.cy/courses/EPL660

1. Introduction
In this session:

- You will learn how to use the ElasticSearch database
- How to index a set of documents
- How to ask simple queries about indexed documents.

2. Running Elasticsearch
Elasticsearch is a distributed, RESTful search and analytics engine capable of solving a growing number of use cases. As the heart of the Elastic Stack, it centrally stores your data so you can discover the expected and uncover the unexpected.

Follow the commands in the accompanying presentation file in order to install Elasticsearch and the related python client libraries.

To test if ElasticSearch is working you have a script called elasticsearch_test.py, you can run it from the command line as:

```
$python elasticsearch_test.py
```

or via Spyder IDE. To open Spyder IDE, call the following command from the terminal:

```
spyder &
```

If Elasticsearch is working you will see an answer from the server, otherwise a message will indicate that it is not running. Alternatively you can check it via the browser at http://localhost:9200

3. Indexing and Querying
As mentioned before, Elasticsearch is accessed as a web service using a REST API. This means that we can operate with the Elasticsearch using the HTTP protocol (just like any web server). Elasticsearch defines a set of methods that correspond to all the actions available. You can access to the full documentation with this link.

To make things simpler we are going to use two python libraries (elasticsearch-py and elasticsearch-dsl) to access Elasticsearch. The first one (elasticsearch-py) is the official low-level client for

Elasticsearch. Its goal is to provide common ground for all Elasticsearch-related code in Python; because of this it tries to be opinion-free and very extendable. It is more general but hides less the complexities of the API calls.

The second library (elasticsearch-dsl) is a higher level client library sitting on top of elasticsearch-py. It is more focused on the search capabilities and is friendlier for sending queries to Elasticsearch. In the scripts that you have in this lab both libraries are used depending on the operations performed.

3.1 Datasets and Indexing

In this lab session's package, we have included two zipped sets of files: 20_newsgroups and novels. Unzip them in your working directory as instructed in the lab8 presentation. They contain respectively:

- Text from 20 usenet groups on various topics, a classic corpus in IR evaluation, from here.
- A number of random novels and other texts in English from the Gutenberg project, with a tendency towards late 19th and early 20th centuries.

You have among the session files a script named IndexFiles.py. Open the script either with Python IDLE or with Spyder (both installed in cloud-based VMs). The script has basically two parts, the first one reads all the documents traversing recursively the path that is received as a parameter and creates a list of Elasticsearch operations. There is a method in the API for indexing just one file, but given that we are going to index many, we can use the bulk method that allows to execute several Elasticsearch calls as one.

A **bulk operation is a special document** that indicates:

(a) the type of operation to execute (`_op_type`), in our case `index`,
(b) the index where the operation is performed (`_index`),
(c) the type of the document (`_type`), in our case `document`,
(d) the fields of the document, in our case we include two fields in the document (1) `path` for the path of the file and (2) `text` for the content of the file.

An example bulk document is shown below:

```
{
    "_op_type": "index",
    "_type": "document",
    "path": "/home/ubuntu/datasets/newsgroups/sci.med/0013590",
    "text": "The contents of the file…"
}
```

The second part of the script connects and interacts with Elasticsearch. Firstly, we need an Elasticsearch object (in our case is called client) that will manage the connection with Elastisearch. Make sure that the Elasticsearch service is working (see presentation of Lab8). Then we can use the command:

```
client = Elasticsearch()
```

given that ElasticSearch service is installed and operating in the local machine with the default configuration. On the other hand, we have to configure the object adequately as shown below:

```python
# you can use RFC-1738 to specify the url
client = Elasticsearch(['https://user:secret@localhost:443'])

# ... or specify common parameters as kwargs
client = Elasticsearch(
    ['localhost', 'otherhost'],
    http_auth=('user', 'secret'),
    scheme="https",
    port=443,
)
```
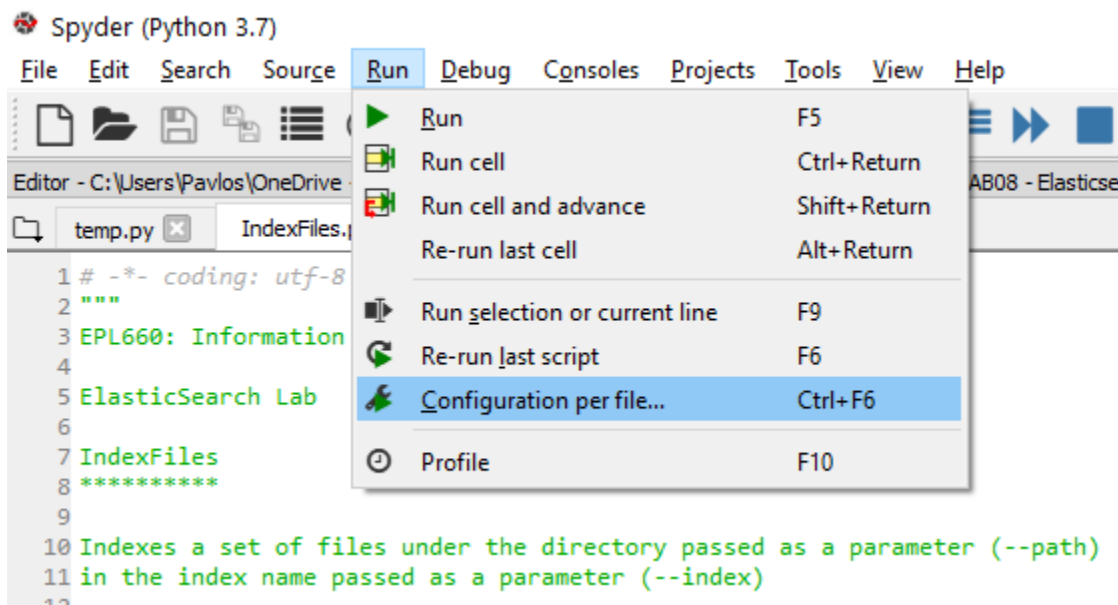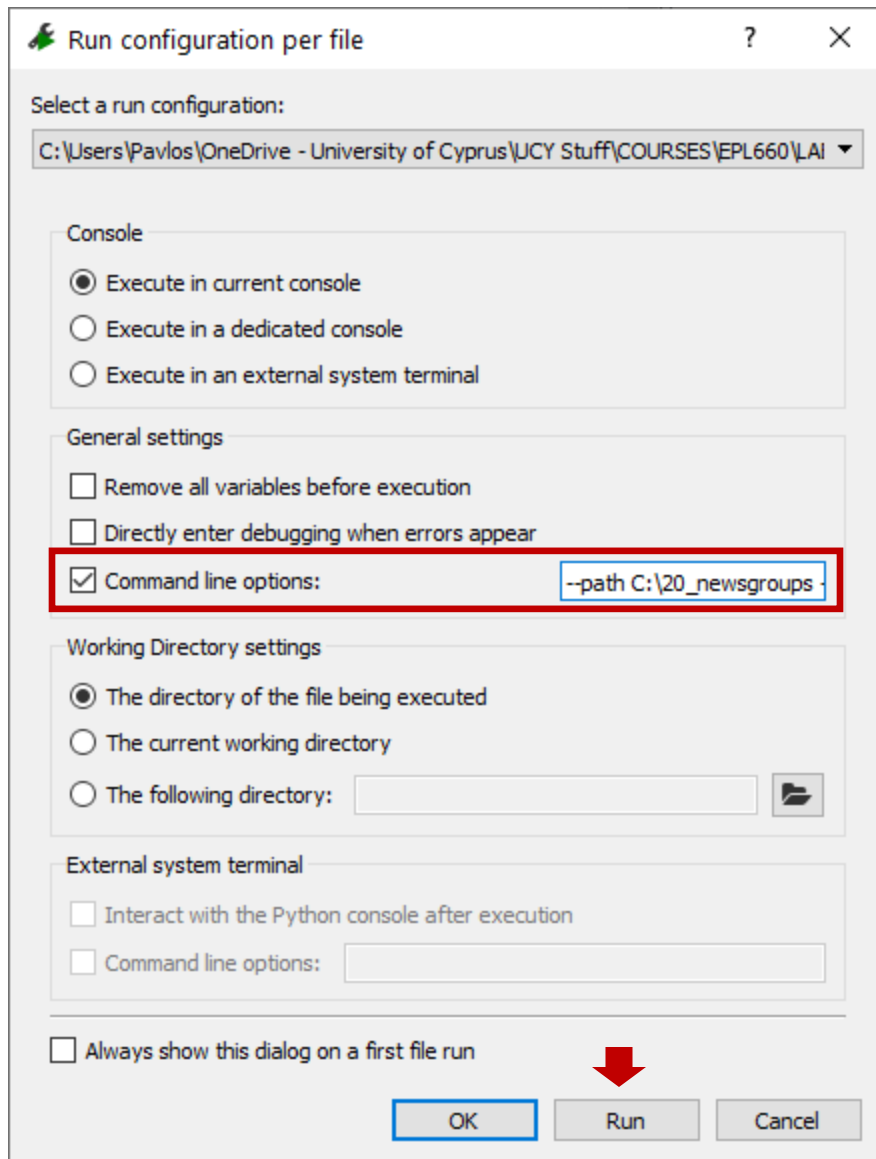
Using the client object, we can create the index for the documents (if it already exists it is deleted) and then the bulk method is called for performing all the document insertions.

Now you can use this script with the documents that you have downloaded. For example, if you run the following command in terminal:

```
$ python IndexFiles.py --path /path/to/20_newsgroups --index news
```

will insert all the documents from 20 newsgroups in the index news. In case you want to run the script within Spyder, please consult the following images on where to place the command line arguments:
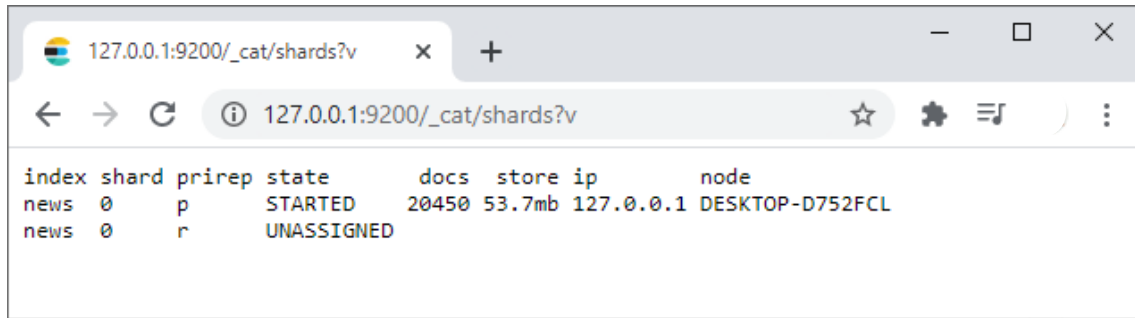
You can check whether the documents were successfully inserted by probing the Elasticsearch REST API via the browser as shown below. As can be seen, the index news expands across 5 primary shards with a total of 20417 documents indexed and stored.
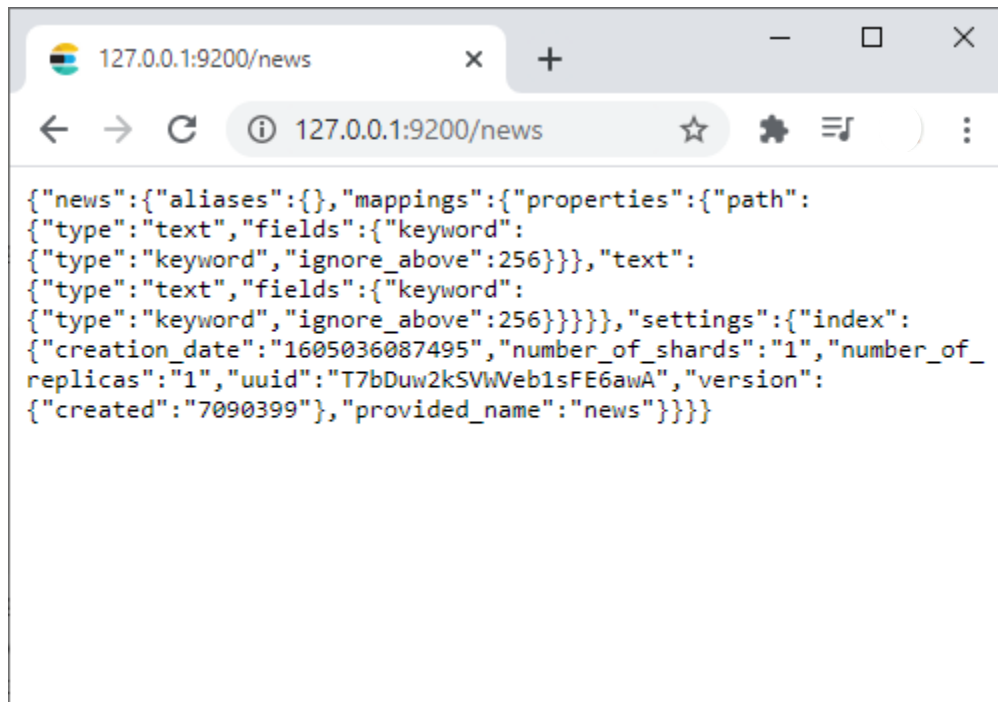

The next screenshot gives more information regarding each shard. As can be observed, the number and size of documents stored on each shard is equally distributed.

127.0.0.1:9200/_cat/shards?v × +

← → C ⓘ 127.0.0.1:9200/_cat/shards?v ☆ ✦ ≡♪) ⋮

```
index shard prirep state       docs  store ip        node
news  0     p      STARTED    20450 53.7mb 127.0.0.1 DESKTOP-D752FCL
news  0     r      UNASSIGNED
```

The next two screenshots the mapping type "document" within the "news" index along with the "path" and "text" properties (fields).

127.0.0.1:9200/news × +

← → C ⓘ 127.0.0.1:9200/news ☆ ✦ ≡♪) ⋮

```
{"news":{"aliases":{},"mappings":{"properties":{"path":
{"type":"text","fields":{"keyword":
{"type":"keyword","ignore_above":256}}},"text":
{"type":"text","fields":{"keyword":
{"type":"keyword","ignore_above":256}}}}},"settings":{"index":
{"creation_date":"1605036087495","number_of_shards":"1","number_of_
replicas":"1","uuid":"T7bDuw2kSVWVeb1sFE6awA","version":
{"created":"7090399"},"provided_name":"news"}}}}
```

3.2 Searching

After having inserted all the documents inside an index we can query ElasticSearch using specific terms or more complicated queries. Elasticsearch provides a JSON-style domain-specific language that you can use to execute queries. This is referred to as the Query DSL. Elasticsearch has a high-level library called Elasticsearch DSL (Domain Specific Language) whose aim is to help with writing and running DSL queries against Elasticsearch. It provides a more convenient and idiomatic way to write and manipulate queries.

In Elasticsearch, apart from exact matches we can do fuzzy matches, aggregations, obtain suggestions, highlight the parts of the document that includes the search term, etc. Basically anything that we expect from a web search engine.

In order to make search requests we will use the Search object from the Search DSL. The Search object represents the entire search request:

- queries
- filters

- aggregations
- sort
- pagination
- additional parameters
- associated client

The script <mark>SearchIndex.py</mark> allows to query an index in ElasticSearch. The given script accepts three arguments, namely `--index` which is obviously the index we have already created as well as the `--text` and the `--query` argument.

The `--text` takes precedence over `--query` (you cannot use both) and allows for searching for a single word in the `text` field of our documents. It will return all the exact matches with the document id, the path of the file that matches and the highlighted parts that contain the word.

The `--query` argument allows to use the LUCENE syntax for the queries. This syntax allows boolean operations like AND, OR, NOT (always in uppercase) or fuzzy searches using `~n` with `n` indicating how many letters of the word can mismatch to consider it a match. Using this argument the query returns the id of the documents, the path and the 20 first characters of the document.

Open the script with a text editor and look to the part of the code where the search query is built. Have a look at the documentation of elasticsearch-dsl to understand it better. Now you can play a little bit with the script, for example execute:

```
$ python SearchIndex.py --index news --text hello
$ python SearchIndex.py --index news --query good AND evil
$ python SearchIndex.py --index news --query angle
$ python SearchIndex.py --index news --query angle~2
```

Each search returns the number of documents matched. Observe how this number changes using larger values for n in fuzzy queries.

Beyond the Python API, you may explore the indexed data by posing queries via the REST API ([search endpoint](#)). For example, you may see the results of the first query in the browser using:

```
http://127.0.0.1:9200/news/ search?q=good
```

as shown below:

127.0.0.1:9200/_search?q=good

{"took":669,"timed_out":false,"_shards":{"total":1,"successful":1,"skipped":0,"failed":0},"hits":{"total": {"value":3854,"relation":"eq"},"max_score":3.4663634,"hits": [{"_index":"news","_type":"_doc","_id":"Bwibs3UB5pDUcux_J72G","_score":3.4663634,"_source": {"path":"C:/20_newsgroups\\misc.forsale/0006202","text":"This is the final public update to my dragon magazine auction. If there\nare no new bids then the current bids stand (like that's gonna happen. :)).\nAfter this, any updates will be by E-mail *ONLY*. The entire auction \nwill end as soon as the bids stop coming in. So if you want to get in\non this, be sure to bid now. All bids must be made in *AT LEAST*\n25 cent increments. Buyer will pay shipping. (Unless you have any\nparticular fancy, it will be US mail 4th class special, with lots of\npadding). \nAll dragons are bagged. The condition of them vary quite a bit, \nso I've come up with my own condition system. Some dragons may be missing \nitems like the inserts. If you have any questions, feel free to ask.\nCondition ratings - Usually this is just an evaluation of the cover, as\nmost of the material inside is in great shape.\nExcellent - IE-\"As you find them in the store\"\nVery Good - Still in great condition, but can tell it's been boughten\nGood   - On down the line\nVery Fair/fair - indicates lots of use (But still \"decent\")\nPoor   - Indicates the material inside may be damaged (usually by\nscissors). \nA * by the condition indicates that something from the magazine is \nmissing (usually the insert)\nI'll post another public update sometime this weekend (which will be the\nfinal public posting, after which the auction will be by e-mail until\nthe bidding stops)\nAgain if you have any questions, ask away.\nnarrow $2.50 056 - Fair   *- Top secret module missing - Bard tunes - \nreal maps\nTN.DE7 $3.0 073 - Good   - Forest of Doom module (detached but \nincluded) - inner planes\nnykchev $1.75 074 - Good   *- Combat computer missing - 4 dragons \nTN.DE7 $1.50 078 - Good/VG   - Monsters - aquatic AD&D module (detached \nbut included) - Language lesson\ngeoffrey $1.50 079 - Good/VG   - top secret module (detached but \nincluded) - magic resistance\nmayla $2.50 081 - Fair   - High level AD&D module (detached but \nincluded) - poison - material spell components\nUCCXKVB $2.00 082 - Very Fair*- Baton races game insert missing - spell \nresearch\nTN.DE7 $1.50 083 - Good/VG   - Babba Yagga's Hut module (detached but \nincluded) - unarmed combat\ngeoffrey $1.50 084 - Fair   *- Cover missing - Twofold talisman module\nsquidly $1.75 085 - Good   - Twofold Talisman module - Clerics\nsquidly $1.75 087 - Good   - Top secret module - Wildernes\ngeoffrey $1.50 088 - Good

If you copy-paste the output of this REST API reply into a JSON validator such as https://jsonlint.com/ you will be able to see the output formatted as shown below:

```
{
    "took": 669,
    "timed_out": false,
    "_shards": {
        "total": 1,
        "successful": 1,
        "skipped": 0,
        "failed": 0
    },
    "hits": {
        "total": {
            "value": 3854,
            "relation": "eq"
        },
        "max_score": 3.4663634,
        "hits": [{
            "_index": "news",
            "_type": "_doc",
            "_id": "Bwibs3UB5pDUcux_J72G",
            "_score": 3.4663634,
            "_source": {
                "path": "C:/20_newsgroups\\misc.forsale/0006202",
                "text": "This is the final public update to my dragon magazine
auction.  If there\nare no new bids then the current bids stand (like that's gonna
happen.  :) ).\nAfter this, any updates will be by E-mail *ONLY*.  The entire auction
\nwill end as soon as the bids stop coming in.  So if you want to get in\non  this,
be sure to bid now.  All bids must be made in *AT LEAST*\n25 cent increments.  Buyer
will pay shipping.  (Unless you have any\nparticular fancy, it will be US mail 4th
```

class special, with lots of\npadding). \nAll dragons are bagged. The condition of them vary quite a bit, \nso I've come up with my own condition system. Some dragons may be missing \nitems like the inserts. If you have any questions, feel free to ask.\nCondition ratings - Usually this is just an evaluation of the cover, as\nmost of the material inside is in great shape.\nExcellent - IE-\"As you find them in the store\"\nVery Good - Still in great condition, but can tell it's been boughten\nGood - On down the line\nVery Fair/fair - indicates lots of use (But still \"decent\")\nPoor - Indicates the material inside may be damaged (usually by\nscissors). \nA * by the condition indicates that something from the magazine is \nmissing (usually the insert)\nI'll post another public update sometime this weekend (which will be the\nfinal public posting, after which the auction will be by e-mail until\nthe bidding stops)\nAgain if you have any questions, ask away.\narrow $2.50 056 - Fair *- Top secret module missing - Bard tunes - \nreal maps\nTN.DE7 $3.0 073 - Good - Forest of Doom module (detached but \nincluded) - inner planes\nykchev $1.75 074 - Good *- Combat computer missing - 4 dragons \nTN.DE7 $1.50 078 - Good/VG - Monsters - aquatic AD&D module (detached \nbut included) - Language lesson\ngeoffrey $1.50 079 - Good/VG - top secret module (detached but \nincluded) - magic resistance\nmayla $2.50 081 - Fair - High level AD&D module (detached but \nincluded) - poison - material spell components\nUCCXKVB $2.00 082 - Very Fair*- Baton races game insert missing - spell \nresearch\nTN.DE7 $1.50 083 - Good/VG - Babba Yagga's Hut module (detached but \nincluded) - unarmed combat\ngeoffrey $1.50 084 - Fair *- Cover missing - Twofold talisman module\nsquidly $1.75 085 - Good - Twofold Talisman module - Clerics\nsquidly $1.75 087 - Good - Top secret module - Wildernes\ngeoffrey $1.50 088 - Good *- Elefant Hunt insert missing - Falling \ndamage - MARVEL-Phile\ngeoffrey $1.50 089 - Good *- Creature catalog missing - Shields - sci fi\nTN.DE7 $1.5 094 - Good - Ranger changes - Creature catalog II \n(detached but included)\ngeoffrey $1.50 095 - Fair *- Cover missing - Into the Forgotten Realms \nmodule, detached but included\nykcheu $2.75 098 - Fair - 9th anniversary - Dragons - mutant manual\nUCCXKVB $1.75 099 - Poor *- Cover Missing - Treasure trove II, some \npictures cut out\nthedm $2.50 100 - Good *- poster missing - city beyond the gate \nmodule (detached but included) - \"raised dragon\" texture on cover\ngeoffrey $1.50 101 - Fair *- Cover missing - creature catalog III \n(detached but included)\ngeoffrey $1.50 102 - Fair *- Cover missing - Valley of earth mother \nmiddle level module (detached but included)\ngeoffrey $1.50 103 - Fair *- Cover loosly attached - Unearth arcana \nupdate missing - Future of AD&D - Centaur papers\ngeoffrey $1.50 103 - Fair *- Unearth arcana update missing \ngeoffrey $1.50 104 - Fair/VF - Marvel module - thieves - cover detached \nbut included \ngeoffrey $1.50 105 - Fair - AD&D module - invisibility - cover & back \ncover detached but included \nUCCXKVB $2.00 106 - Fair - Cover 1/2 on - variations of paladins - \nmore skills 4 rangers\nykcheu $2.75 106 - Good/VG - Variations of paladins - more skills for \nrangers \narrow $2.50 107 - Fair *- Cover missing - Dragons of glory \nsupplement/questionaire\ngeoffrey $1.50 108 - Good - Mutant manual II - environmental \neffects - cover taped reinforced\nthedm $2.00 108 - Very Good - Mutant manual II - environmental effects \nykcheu $2.25 109 - Good - Customizing D&D classes - Agent 13 poster\ngeoffrey $1.50 109 - Very Fair - Customizing D&D classes - Agent 13 poster \nmissing\ngeoffrey $1.50 110 - Very Good - House on the frozen lands module -10th anniv\nsquidly $1.75 110 - Very Good - House on the frozen lands module -10th anniv\ngeoffrey $1.50 111 - Good - Murder Mystery AD&D module\nykcheu $2.50 112 - Very Good - Ultimate Article Index - Mesozoic monsters\nmayla $2.50 114 - Very Fair - Elven Cavalier - remorhaz - Witch NPC class\nykcheu $1.75 115 - Good - Theives - harpies & snakes\nsquidly $1.75 116 - Good/VG - 3-D ship

cardboard insert - wild \nanimals - dr who\nUCCXKVB $2.50 117 - Good/VG   - Dice odds - creative campaigns - sage \nadvice - bazaar\ngeoffrey $1.50 118 - Good   - Tournaments/Competitions - Nibar's keep game\nUCCXKVB $2.00 120 - VG/EX   - April fool's issue\nUCCXKVB $2.00 121 - Excellent - Oriental adventures - cardboard castle \ninsert\ngeoffrey $1.50 122 - Excellent - 11th aniversary - African beasts - druids\nUCCXKVB $2.50 123 - Very Good - Magic and wizardry\nthedm $2.25 123 - Very Good - Magic and wizardry\narrow $2.5 124 - Excellent - Aerial adventures - 2nd edition ?aire\nUCCXKVB $2.25 124 - Excellent - Aerial adventures - 2nd edition ?aire\ngeoffrey $1.50 125 - Very Good - Clay-O-Rama! - Chivalry - quasi elementals\ngeoffrey $1.50 125 - VG/EX   - Clay-O-Rama! - Chivalry - Quasi-elementals\nUCCXKVB $2.50 126 - VG/EX   - Undead\nUCCXKVB $2.00 127 - Very Good - Fighters\n2FVPMANTEL $3.00 128 - Good   - King's Table insert game\n2FVPMANTEL $3.00 129 - Excellent - Demi-humans\nUCCXKVB $2.50 130 - VG/EX - The arcane arts\nTFPAYN01 $2.50 131 - Excellent - Deepearth\narrow $1.5 131 - VG/EX   - Deepearth\nUCCXKVB $2.0 132 - Very Good - ORCWARS! board game missing\narrow $1.5 133 - Very Good - Berserkers & Spies-Roman gods -marvel index\ngeoffrey $1.50 133 - Very Good - Berserkers & Spies-Roman gods -marvel index\n2FVPMANTEL $3.00 134 - VG/EX   - 12 anniversary - Dragons \ntbh1 $2   135 - Very Good - Archers - Space sage advice\nUCCXKVB $2.0 135 - Very Good - Archers - Space Sage advice\ntbh1 $4.0 136 - Very Good - Cities & Urban adventures\ntbh1 $3   137 - Excellent - Wilderness\narrow $1.5 138 - Very Good - Horror (Haloween)\nUCCXKVB $3.25 139 - Very Good - Pages from the Mages\nUCCXKVB $3.5 140 - Excellent - Clerics & Healers\ntbh1 $3   141 - Good   - Humanoids\ntbh1 $2   142 - Very Good - AD&D 2nd edition preview\nUCCXKVB $3.25 143 - Very Good - DM's issue\nUCCXKVB $2.00 145 - Very Good*- Poster missing - castles\nthedm $2.5 146 - Very Good*- Poster missing - 13 anninversary - Dragons\ntbh1 $3.00 147 - Excellent - MAGUS! board game - magic\ntbh1 $3.5 148 - Excellent - Fighting - Deck of Many things insert\nTN.DE7 $2.0 149 - Excellent - (No particular feature)\nTN.DE7 $2.00 150 - Excellent - Horror (Halloween issue)\nTN.DE7 $2.00 151 - Excellent - Oriental Adventures/Eastern \nTN.DE7 $2.00 152 - Good   *- Underdark - Poster missing is inside-\nslight crumple on cover, only noticable under inspection\nTN.DE7 $2.0 153 - Very Good - Gods\nTN.DE7 $1.75 154 - Good/VG - Poster - Dragonlance story - War\nTN.DE7 $2.00 155 - Excellent - Faeries - DUNGEON module\nTN.DE7 $1.75 157 - Very Good - Buck Rogers\nthedm $2.00 158 - Very Good - 14th anniversary - Dragons\nUCCXKVB $2.00 159 - Excellent - Spelljammer - Poster missing\nkohlmaas $2.00 160 - Good/VG *- Urban adventures - AD&D trading card \ninsert missing\nUCCXKVB $2.25 161 - Very Good - DM issue\nTN.DE7 $1.75 162 - Good   - Haloween - Poster missing\nthedm $2.0 163 - Excellent - Monsterous compendum insert - Magic\nTN.DE7 $2.0 164 - Very Good - Oriental Adventures\nTN.DE7 $1.75 165 - VG/EX   - Sea/Undersea\nTN.DE7 $1.75 166 - Excellent*- Sci Fi (other games) - Dino wars insert \nmissing\nTN.DE7 $1.75 167 - Excellent - Nature/Wilderness\nUCCXKVB $2.5 169 - Very Good - Slight crease of back cover - Misc \nitems featured\nthedm $2.5 170 - Good   - Slight crease in cover - Dragon kings game \ninsert - Dragons - 15th anniversary issue\nUCCXKVB $2.00 171 - Excellent*- Missing poster & trading cards (ARGH!)\nCfrye $2.75 172 - Excellent - Underdark\nhachiman $2   173 - Excellent - Dark Sun\nintravai $2   174 - Excellent - Horror\nintravai $3.00 175 - Excellent - World building - Campaign help\nTN.DE7 $2.00 176 - Excellent - Elves - Giant poster inside\nTN.DE7 $1.5 177 - Very Good - Calender poster - DM help (gunpowder too)\nintravai $3.0 178 - Excellent - Fighters & the Fighter class\nTN.DE7 $2.50 179 - Excellent - GENCON form - Magic items featured\nTN.DE7 $2.50 181 - Excellent - Calendar Poster - Mages/Sorcerors\nTN.DE7 $2.50 182 - Excellent - 16 anniversary issue - Dragons\nTN.DE7 $2.50 184 - Excellent - Non Player Character enhancement\nTN.DE7

$1.75 185 - Excellent - Dark Sun Campaign Monsters - Dark Sun\ngeoffrey $1.50 186
- Excellent - Haloween - Horror\nTN.DE7 $1.75 187 - Excellent - Wilderness -
Outdoors\nIf you notice any errors, please let me know (other than slight name
\nmisspellings, if it's close to your name, that's you. :))\n"
                }
        }, {
                "_index": "news",
                "_type": "_doc",
                "_id": "Hgibs3UB5pDUcux_GqaL",
                "_score": 3.3158636,
                "_source": {
                        "path": "C:/20_newsgroups\\alt.atheism/0000388",
                        "text":   "In   article
<1993Apr13.173100.29861@ultb.isc.rit.edu> snm6394@ultb.isc.rit.edu (S.N. Mozumder)
writes:\n>>I'm only saying that anything can happen under atheism. Being
a\n>>beleiver, a knowledgeable one in religion, only good can happen.\nThis is
becoming a tiresome statement. Coming from you it is \na definition, not an
assertion:\n   Islam is good. Belief in Islam is good. Therefore, being a \n
believer in Islam can produce only good...because Islam is\n   good. Blah blah
blah.\nThat's about as circular as it gets, and equally meaningless. To\nsay that
something produces only good because it is only good that \nit produces is nothing
more than an unapplied definition. And\nall you're application is saying that it's
true if you really \nbelieve it's true. That's silly.\nConversely, you say off-
handedly that _anything_ can happen under\natheism. Again, just an offshoot of
believe-it-and-it-becomes-true-\ndon't-believe-it-and-it-doesn't.   \nLike other
religions I'm aquainted with, Islam teaches exclusion and\ncaste, and suggests harsh
penalties for _behaviors_ that have no\nlogical call for punishment (certain limits
on speech and sex, for\nexample). To me this is not good. I see much pain and
suffering\nwithout any justification, except for the _waving of the hand_ of\nsome
inaccessible god.\nBy the by, you toss around the word knowledgable a bit
carelessly.\nFor what is a _knowledgeable believer_ except a contradiction of\nterms.
I infer that you mean believer in terms of having faith.\nAnd If you need knowledge
to believe then faith has nothing\nto do with it, does it?\n-jim halat\n"
                }
        }, {
                "_index": "news",
                "_type": "_doc",
                "_id": "2wibs3UB5pDUcux_NMjO",
                "_score": 3.315724,
                "_source": {
                        "path": "C:/20_newsgroups\\rec.sport.baseball/0009187",
                        "text":   "In   article   120399@netnews.upenn.edu,
sepinwal@mail.sas.upenn.edu (Alan Sepinwall) writes:\n>And now, the not so pleasant
surprises:\n>\t2)Tartabull. The book on Tartabull was, keep him healthy and\n>\t
he'll produce. Well, he hasn't done too much so far. Sure. he's\n>\t hit a few
homers, but those were all solo shots, and he hasn't\n>\t gotten any of the \"big\"
RBIs that your cleanup man is supposed\n>\t to give you. Then again, he had a slow
start last year (once he\n>\t got off the DL, that is) and turned into a one-man
wrecking crew\n>See you in the Series!\n>-Alan\nLet's see... April 15th... less than
30 at bats.... and you claim that he \nhasn't done too much so far!\nCut this guy
some slack. Danny will produce this year. It's scary to think\njust how much he'll
produce if he were to stay healthy all year.\nThe Yanks have a lot going for them
this year: good starting rotation, good\nbullpen, good defense and a good lineup.
Also, I like Buck Showalter. Frank\nHoward on 1st is also a good move. Everything

```
sounds good so far. \nIf the Yanks stay healthy, they have a good chance at winning
the pennant. This \nis the most fun I've had watching the Yanks since \"78!\n-
Bobby\n"
                }
        }, {
                "_index": "news",
                "_type": "_doc",
                "_id": "yAibs3UB5pDUcux_GqWL",
                "_score": 3.3155375,
                "_source": {
                        "path": "C:/20_newsgroups\\alt.atheism/0000302",
                        "text":          "In           <kmr4.1576.734879396@po.CWRU.edu>
kmr4@po.CWRU.edu      (Keith       M.      Ryan)      writes:\n>In      article
<1qj9gq$mg7@horus.ap.mchp.sni.de>      frank@D012S658.uucp      (Frank      \nO'Dwyer)
writes:\n>>Is good logic *better* than bad?  Is good science better than bad?  \n>
By definition.\n> great - good - okay - bad - horrible\n>   << better\n>       worse
>>\n> Good is defined as being better than bad.\nHow do we come up with this setup?
Is this subjective, if enough people agreed\nwe could switch the order?  Isn't this
defining one unknown thing by another? \nThat is, good is that which is better than
bad, and bad is that which is worse\nthan good?  Circular?\n>   Only when the Sun
starts    to    orbit    the    Earth    will    I    accept    the    Bible.    \n
Michael A. Cobb\n \"...and I won't raise taxes on the middle      University of
Illinois\n    class to pay for my programs.\"                    Champaign-Urbana\n
-Bill Clinton 3rd Debate            cobb@alexia.lis.uiuc.edu\nWith new taxes and
spending cuts we'll still have 310 billion dollar deficits.\n"
                }
        }, {
                "_index": "news",
                "_type": "_doc",
                "_id": "oAibs3UB5pDUcux_OtQe",
                "_score": 3.298139,
                "_source": {
                        "path": "C:/20_newsgroups\\sci.electronics/0012192",
                        "text": "davidj@rahul.net (David Josephson) writes:\n>In
<C5JJJ2.1tF@cmcl2.nyu.edu> ali@cns.nyu.edu (Alan Macaluso) writes:\n>>I'm looking to
build a microphone preamp that has very good low-noise characteristics,  large clean
gain, and incorportates phantom power (20-48 volts (dc)) for a PZM microphone.  I'm
leaning towards a good, low-cost (??) instrumentation amplifier to maintain the
balanced input from the microphone, for its good CMRR, internal compensation, and
because i can use a minimal # of parts.  \n>>Does anyone out there have any experience,
suggestions, advice, etc...that they'd like to pass on, I'd greatly appreciate
it.\n>Without doing anything really tricky, the best I've seen is the\n>Burr-Brown
INA103. Their databook shows a good application of this\n>chip as a phantom power
mic pre.\nI've had very good results from the SSM2016 from PMI (part of
Analogue\nDevices). They have also now introduced the SSM2017 which looks good
on\npaper, but which I haven't tried yet.\nChristopher\n  Christopher Hicks     |
Paradise is a Linear Gaussian World\n  cmh@uk.ac.cam.eng     |    (also reported to
taste hot and sweaty)\n"
                }
        }, {
                "_index": "news",
                "_type": "_doc",
                "_id": "Mwibs3UB5pDUcux_NMZc",
                "_score": 3.2632527,
```

                    "_source": {
                            "path": "C:/20_newsgroups\\rec.motorcycles/0008507",
                            "text": "One point not yet mentioned:\n\tHands on the driver's shoulders are a definite no-no. It feels\ngood for the passenger to stretch her arms, and a shoulder massage feels\ngood too, but the shoulders are connected to the arms, and the arms to\nthe handlebars... Should you hit the brakes, the pressure on the shoulders\ncan make steering interesting.\nBundle her up if it's cold, make sure she has good gear (say it again),\nan pat her thigh to let her know all is great.\nGood luck!\t\tMike\n"
                    }
            }, {
                    "_index": "news",
                    "_type": "_doc",
                    "_id": "-gibs3UB5pDUcux_J7yG",
                    "_score": 3.23341,
                    "_source": {
                            "path": "C:/20_newsgroups\\misc.forsale/0006189",
                            "text":   "Hi,\nI have the following software forsale:\nMicrosoft   MS-DOS 5.0   3.5\"DD $15\n-- This is a good buy for those who don't need all the utils in DOS 6.0\nAccolade   Hardball II   5.25\"DD $10\n-- Good arcade baseball game, graphics/sound are pretty good, has the\n   ability to make schedules and edit player stats\nAccolade The Third Courier   5.25\"HD $5\n-- Adventure/Spy type game\nEGA Earl Weaver 2   3.5\"DD $15\n-- Good arcade/strategy type baseball game, you have ability to play\n   for league play, i have MLB stats for 1990 and 1991\nEGA   Stormovik SU25: Soviet Attack Fighter 3.5\"DD $10\n-- Good fighter simulation, various missions and levels of play\n Sierra   Thexder FireHawk   3.5\" & 5.25\" $5\n-- Arcade shoot 'em up type game, nice music and sound effects\nVirgen   Scrabble   3.5\" $5\n-- computer version of the popular board game\nI'm willing to negotiate on the prices, prices do not include shipping.\nShipping costs will be split 50/50. Drop me a line! :)\nRoberto Alvarez   ud156844@ndsuvm1\nProgrammer Analyst   ud156844@vm1.nodak.edu\nUniversity of North Dakota   adrxa@undjes2\n"
                    }
            }, {
                    "_index": "news",
                    "_type": "_doc",
                    "_id": "Ngibs3UB5pDUcux_M8Tl",
                    "_score": 3.213571,
                    "_source": {
                            "path": "C:/20_newsgroups\\rec.autos/0007999",
                            "text": "In article <1993Apr27.172335.23392@magnus.acs.ohio-state.edu> bbarnard@magnus.acs.ohio-state.edu (Bret D Barnard) writes:\n>With Spring upon us it's time to start waxing again. I was wondering if anyone\n>had any suggestions about waxes and polishes. I know carnuba wax is often \n>recommended on the net, but I was wondering about the colored waxes to cover up\n>some minor scratches. Do the colored waxes give on overall good quality shine?\n>Are they good for the paint or clear coat?\n>What about Liquid Glass, is it a clear polish which dries clear; thus filling\n>scratches? is it good for the paint/clear coat?\n>Please post with both good and bad experiences. And hey, let's name names \n>(brands) here! ;^)\nAll gimmicks! Stick with plain 'ol carnauba wax that's non-abrasive.\nEagle 1, Meguiars, Turtle Wax, and a few others are good examples.\nThe colored waxes just color in the scratches so they're not so \napparent. The better approach is to buff

minor scratches completely\noff with a cleaner/mild abrasive.\nNever tried Liquid Glass, although I still have this sample they sent\nme a few years back. \naaron\n"
                    }
            }, {
                    "_index": "news",
                    "_type": "_doc",
                    "_id": "Cwibs3UB5pDUcux_IazS",
                    "_score": 3.2009304,
                    "_source": {
                            "path": "C:/20_newsgroups\\comp.graphics/0001889",
                            "text": "In article 1rfsqbINNc2p@shelley.u.washington.edu, stusoft@hardy.u.washington.edu (Stuart Denman) writes:\n>Does anyone out there have any JPEG decompression code in pretty much any\n>language that I can read and understand? I have trouble understanding the\n>JPEG Group's code that I got from an FTP site. If any one can send me\n>some good code, I will appreciate it a lot! Thanks!\nThe problem is that the process is inherently complicated ! The IJG's code is\npretty good if you ask me, and I have watched it go through many many cycles of\nrevision.\nTry getting a good book on the subject, that will explain the algorithms.\nSpecifically \"JPEG Still Image Compression Standard\" by Pennebaker & Mitchell,\nBTW. I presume your comment about \"good\" code wasn't meant to sound as offensive\nas it does.\nDavid A. Clunie (dclunie@pax.tpa.com.au)\n"
                    }
            }, {
                    "_index": "news",
                    "_type": "_doc",
                    "_id": "Mgibs3UB5pDUcux_Otaz",
                    "_score": 3.2009304,
                    "_source": {
                            "path": "C:/20_newsgroups\\sci.electronics/0012594",
                            "text": "klink@cbnewsl.cb.att.com   (steven.r.klinkner) writes:\n> Can anybody recommend a good, application-oriented beginner's reference\n> to RF circuits? \n> I am pretty good on theory & know what different types of modulation mean, \n> but don't have a lot of practical experience. A book detailing working\n> circuits of different types (modulation, power, frequency, what is legal,\n> what is not, et cetera), would be very helpful.\n> Thanks.\nWell, you might try the A.R.R.L.'s license study guides. For example, my\nAdvanced Class study guide has lots and lots of good RF and electronics\ntheory in it. I would imagine the other books are good too.\nTony\n-- Anthony S. Pelliccio, kd1nr/ae   // Yes, you read it right, the //\n-- system @ garlic.sbs.com   // man who went from No-Code //\n--------------------------------// (Thhhppptt!) to Extra in   //\n-- Flame Retardent Sysadmin   // exactly one year!   //\n-- This is a calm .sig! --\n"
                    }
            }]
        }
}

Submission

When you finish all the steps of this lab, create a word document containing the results of the following queries using both the SearchIndex.py (copy text or screenshot from Spyder console) and the REST API (screenshot from browser):

patriot OR citizen

angle AND distance