Search Engines (indexing)

검색엔진 입문에서 응용까지

1) http://www.comworld.co.kr/news/articleView.html?idxno=49504
2) http://www.comworld.co.kr/news/articleView.html?idxno=49520
3) http://www.itdaily.kr/news/articleView.html?idxno=91570
4) http://www.comworld.co.kr/news/articleView.html?idxno=49575
5) http://www.comworld.co.kr/news/articleView.html?idxno=49593
6) http://www.itdaily.kr/news/articleView.html?idxno=93371
7) http://www.comworld.co.kr/news/articleView.html?idxno=49626

내 맘대로 es 정리 중

https://github.com/itmare/es_lecture

ElasticSearch 를 이용한 PDF 와 Word 문서 검색 서비스 만들기

https://naggingmachine.tistory.com/830

elastic search 제대로 배워보자

https://www.oss.kr/storage/app/public/oss/87/ea/[Elasticsearch]%20Solution%20Guide%20V0.95.pdf

Elasticsearch 6.x Cheatsheet

https://elasticsearch-cheatsheet.jolicode.com/#es6

Apache Solr vs Elasticsearch

https://solr-vs-elasticsearch.com/

엘라스틱서치 편의 기능

Building a real-time elastic search engine using Python

https://medium.com/faun/building-a-real-time-elastic-search-engine-using-python-32e05bcb9140

Crawling - PDF

Blog - Python for PDF
https://towardsdatascience.com/python-for-pdf-ef0fac2808b0
PDFMiner (manual)
https://pdfminersix.readthedocs.io/en/develop/
PDFMiner (github)
https://github.com/pdfminer/pdfminer.six
tabula-py: Extract table from PDF into Python DataFrame
https://blog.chezo.uno/tabula-py-extract-table-from-pdf-into-python-dataframe-6c7acfa5f302
Camelot: PDF Table Extraction for Humans
https://camelot-py.readthedocs.io/en/master/
Extract references (pdf, url, doi, arxiv) and metadata from a PDF; optionally download all referenced PDFs http://www.metachris.com/pdfx

Ranking

"선형대수학과 구글(Google) 검색엔진" - 페이지랭크 알고리즘
http://matrix.skku.ac.kr/2012-e-Books/KMS-News-LA-Google-SGLee.pdf

[텍스트마이닝] 구글 페이지랭크(PageRank) 알고리듬
http://www.emh.co.kr/content.pl?google_pagerank_citation_ranking

한글 텍스트 전처리
https://programmers.co.kr/learn/courses/21/lessons/1694

추천 using collaborative filtering
https://proinlab.com/archives/2103

Natural Language Processing

한국어 분석(형태소 분석)

https://hoony-gunputer.tistory.com/79

Spacy

https://spacy.io/

Topic Models

Gensim

https://radimrehurek.com/gensim/

SE using Model Models

벡터 필드를 사용한 텍스트 유사도 검색

https://www.elastic.co/kr/blog/text-similarity-search-with-vectors-in-elasticsearch?fbclid=IwAR21aBmhTr6C1ZeNGpxVsLubNa90MjeZ81nU3YGzKqnKuV1xJUlmRjQYMwg