

Lucene-6.x 기초 및 활용 강좌



영플러스소프트는 Lucene/Solr 기반의 검색 엔진(ZoeSearch) 및
Big Data 분석,처리 전문업체로,
다양한 콘텐츠 검색과 서비스를 위한 콘텐츠 자동 추출을 통해
고객서비스의 Quality 향상을 지원해 드리고 있습니다.

- 회 사 명 : 주식회사 영플러스소프트 (대표:강명구)
- 설립연도 : 2011년 01월
- 사업분야 : 검색엔진 솔루션 및 Big Data 분석, 처리 기반 사업
- 본부주소 : 서울특별시 송파구 가락로 71 332호
- 개발본부주소 : 경기도 성남시 수정구 수정남로 276번지 2층
- 연 락 처 : TEL. 02-417-2251 / 031-735-4920, FAX. 031-735-4919
<http://www.youngplussoft.com> / mgkaki@youngplussoft.com

YPS의 파트너 & 고객

Client



Lucene/Solr 검색엔진의 특징

기술적인 특징들을 기술하였습니다.

□ 검색 엔진 및 솔루션

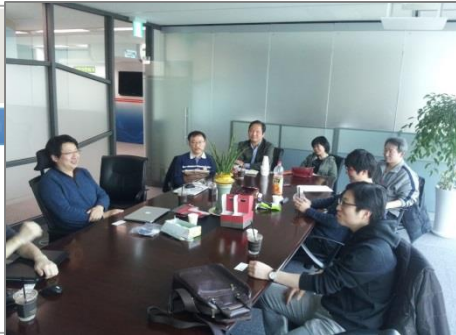
- ZoeSearch(Lucene/Solr 기반 영플러스소프트 검색엔진) 의 특징
 - 다국어 형태소 분석기 : 한국어, 영어, 중국어, 일본어
 - 다양한 문서 파서 보유 : HWP, MS-Office, AutoCAD, PDF....
 - Spatial Search 기능 (ex: 지도 상의 위치 기반 검색 가능)
 - BigData 플랫폼과 연동 가능(Hadoop & Mahout)
 - 분산 색인 및 이중화가 지원되는 대용량 검색 기능
 - Fortune 지 선정 100대기업에서 사용중인 안정적인 오픈 소스 검색엔진
 - 국내에서도 특허정보원 및 대기업에서 상용으로 사용 중
 - 전세계 Open-Source 그룹 중 가장 활발하게 활동하는 커뮤니티
 - 신기술의 변화가 가장 빨리 적용됨

Lucene/Solr 커뮤니티

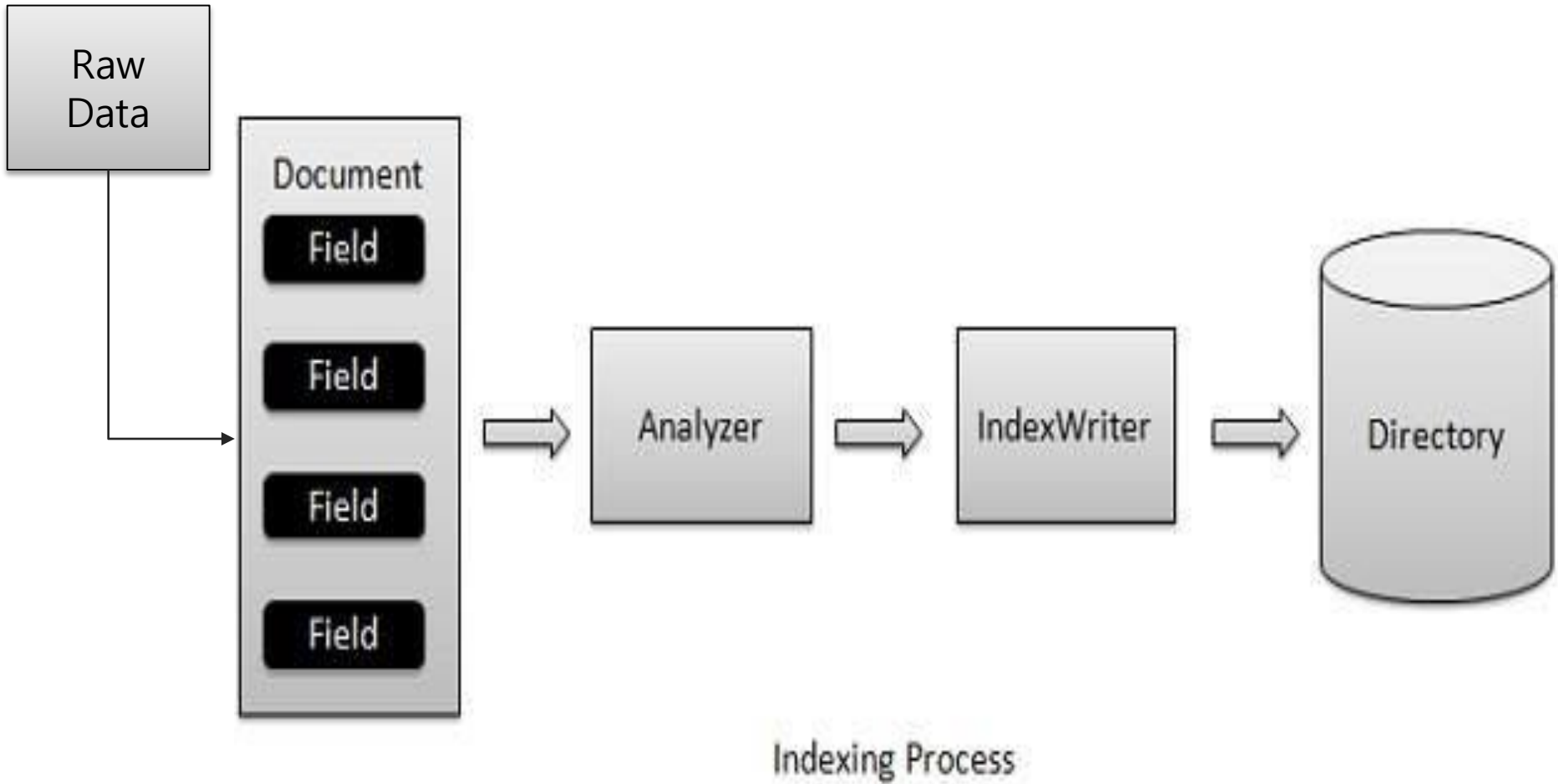
실제 적용된 사업 및 대외 활동에 대하여 기술합니다.

❑ Lucene/Solr 검색엔진 국내 커뮤니티

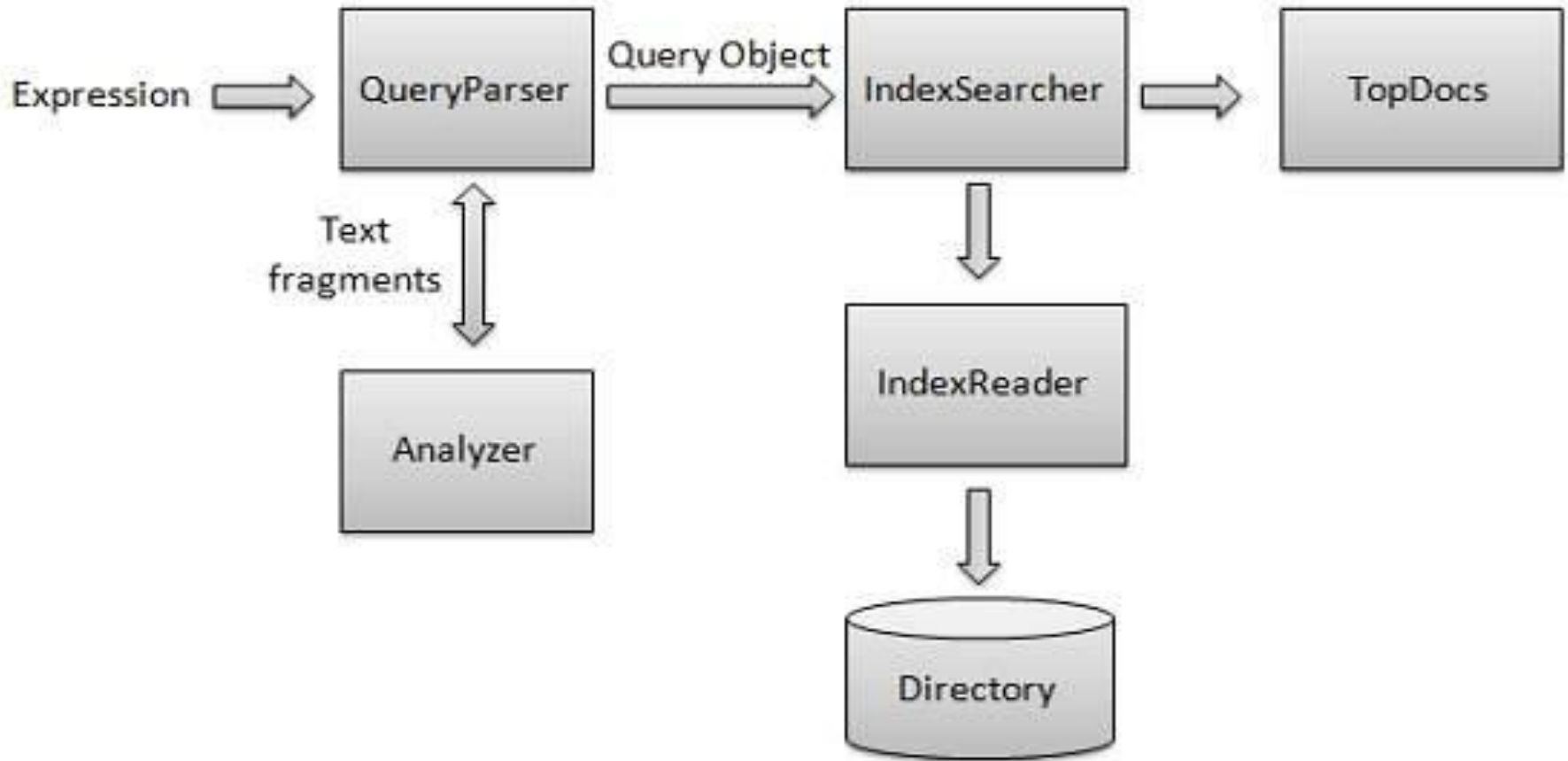
- Lucene/Solr Community : <http://cafe.naver.com/korlucene>
- 연구모임을 결성하여 기술세미나 실시, 개발 경험 공유 및 상호지원



Lucene Indexing 과정



Lucene Search 과정



Searching Process

용어 정의

Database	Lucene	설명
Table	Index	검색 색인
Column	Field	색인 필드
Row	Document	색인 문서 단위 데이터
Sql	Term	한 구절(word)를 나타냄
	Query	색인 질의 오브젝트
	Analyzer	색인 & 검색 언어 분석기
	QueryParser	색인 질의를 위한 파서 (Analyzer 부터 생성)
	Directory	색인 디렉토리 관리기
	Reader	색인 Reader (Directory 로 부터 생성)
	Searcher	색인 검색기 (Reader 로 부터 생성)
driver	IndexWriter	색인 생성기 (Directory & Analyzer 로 부터 생성)
insert	addDocument	IndexWriter.addDocument(document) 도큐먼트 추가
update	updateDocument	IndexWriter.updateDocument(term, document) 도큐먼트 수정
delete	deleteDocument	IndexWriter.deleteDocument(delTerm, document) 도큐먼트 삭제
PrimaryKey	uniqueId	도큐먼트의 고유 ID (추상적인 필드)
commit	commit	데이터 업데이트 commit

Index

Document1

Field A (name/value)

Field B (name/value)

Document2

Field A (name/value)

Field B (name/value)

....

- ❑ `FSDirectory.open(Path path)`
 - 일반적인 디렉토리
- ❑ `NIOFSDirectory(Path path[, LockFactory lockFactory])`
 - 멀티 스레드 환경에서 보다 안정적인 NIOFS 지원 디렉토리
- ❑ `RAMDirectory([LockFactory lockFactory])`
 - 메모리 기반 디렉토리(like memoy database table)
- ❑ `RAMDirectory(FSDirectory dir, IOContext context)`
- ❑ `IOContext` : Merge 조건등 설정 : 최대 문서 수, 최대 머즈 바이트, 최대 머즈 세그먼트 수...

```
Path path = FileSystems.getDefault().getPath(dir);
```

```
Directory directory = new NIOFSDirectory(path) ;
```

```
//Directory directory = new FSDirectory.open(path) ;
```

```
IndexWriterConfig config = new IndexWriterConfig(analyzer) ;
```

```
IndexWriter writer = new IndexWriter(directory , config) ;
```

Lucene Document Field 종류

- ❑ StringField : 저장용 문자열 필드
 - Range 필터링 필드로도 사용 가능
- ❑ TextField : fulltext 검색용 문자열 필드
- ❑ NumericField : **IntPoint, LongPoint, FloatPoint, DoublePoint**
 - Range 필터링 용도로 많이 사용되는 필드
- ❑ DocValuesField : Scoring & Sorting에 효율적인 필드.
 - ByteDocValueField, DerefBytesValuesField, DoubleDocValuesField, NumericDocValuesField, PackedLongDocValuesField, ShortDocValuesField, SortedBytesDocValuesField, StraightBytesDocValuesField,

Lucene Document 생성

```
Document doc = new Document();
```

```
doc.add(new StringField("id", "1" , Field.Store.YES));
```

```
doc.add(new StringField("name", "르네상스 호텔", Field.Store.YES));
```

```
doc.add(new StringField("city", "서울", Field.Store.YES));
```

```
doc.add(new IntPoint("members", 1234)); // Field.Store.NO
```

```
doc.add(new StoredField("members", 1234)); //Field.Store.YES
```

```
String fullTextDesc = "르네상스 호텔은 서울에 위치하고 있으면  
최상위 층에 최고 시설의 스위트 룸을 구비하고 있습니다. 총 직원은  
1234 명이고....." ;
```

```
Doc.add(new TextField("content", fullTextDesc, Field.Store.NO));
```

Add Document

```
IndexWriter writer = getIndexWriter() ;
Document doc = new Document();
doc.add(new StringField("id", "1" , Field.Store.YES));
doc.add(new StringField("name", "르네상스 호텔", Field.Store.YES));
doc.add(new StringField("city", "서울", Field.Store.YES));
doc.add(new IntPoint("members", 1234)); // Field.Store.NO
doc.add(new StoredField("members", 1234)); //Field.Store.YES
String fullTextDesc = "르네상스 호텔은 서울에 위치하고 있으면
최상위 층에 최고 시설의 스위트 룸을 구비하고 있습니다. 총 직원은
1234 명이고....." ;
doc.add(new TextField("content", fullTextDesc, Field.Store.NO));
writer.addDocument(doc) ;
```

Update Document

```
IndexWriter writer = getIndexWriter() ;
Document doc = new Document();
doc.add(new StringField("id", "1" , Field.Store.YES));
doc.add(new StringField("name", "르네상스 호텔", Field.Store.YES));
doc.add(new StringField("city", "부산", Field.Store.YES));
doc.add(new IntPoint("members", 1234)); // Field.Store.NO
doc.add(new StoredField("members", 1234)); //Field.Store.YES
String fullTextDesc = "르네상스 호텔은 부산에 위치하고 있으면
최상위 층에 최고 시설의 스위트 룸을 구비하고 있습니다. 총 직원은
1234 명이고....." ;
doc.add(new TextField("content", fullTextDesc, Field.Store.NO));
writer.updateDocument(new Term("id", "1"), doc) ;
```

deleteDocument

- ❑ `writer.deleteDocuments(new Term("id", "1")) ;`
 - Term 으로 매칭되는 도큐먼트 모두 삭제
- ❑ `writer.deleteDocuments(query) ;`
 - Query 생성에 대해서는 뒷 페이지에서 설명
 - Query 검색 결과에 해당하는 모든 도큐먼트 삭제


```
String indexPath = "./index" ;  
Path path = FileSystems.getDefault().getPath(indexPath);  
Directory directory = new NIOFSDirectory(path);  
IndexReader reader = DirectoryReader.open(directory);  
IndexSearcher searcher = new IndexSearcher(reader) ;
```

indexPath 에 생성되어 있는 색인을 검색하기 위한 Searcher 생성
(Open) 순서는 Directory -> IndexReader -> IndexSearcher 순이다.

Query Parser 를 이용한 Query 생성

```
Analyzer analyzer = new KoreanAnalyzer(version);
```

- 아리랑 한글 분석기 생성

```
QueryParser qp = new QueryParser("content", analyzer);
```

- "content" 필드에 대한 한글 분석기 쿼리 파서 생성

```
Query query = qp.parse("서울에 있는 호텔");
```

- "서울에 있는 호텔" 질의어 분석 Query 생성

TermQuery 생성

- ❑ Query query = new TermQuery(new Term("city", "서울"));
 - City 필드가 "서울" 인 문서를 찾기 위한 Query

- ❑ Query rquery =
 new TermRangeQuery("pubdate",
 new BytesRef("20140101".getBytes()),
 new BytesRef("20150101".getBytes()), true, false);
 - true[시작/끝 range 에 포함], false[시작/끝 range 에 미포함]

NumericRangeQuery

Query query =

IntPoint.newRangeQuery("members", 100, 1000);

- “members” 필드의 값이 100 ~ 1000인 도큐먼트를 검색하기 위한 RangeQuery

Boolean Query

```
QueryParser qp = new QueryParser("content", analyzer);  
Query pquery = qp.parse("서울에 있는 호텔");
```

```
Query nquery =  
    IntPoint.newRangeQuery("members", 100, 1000);
```

```
Builder builder = new Builder() ;  
//FILTER : AND(Not Scoring), MUST : AND, SHOULD : OR, MUST_NOT : NOT -  
Boolean Query  
builder.add(pquery, BooleanClause.Occur.MUST);  
builder.add(nquery, BooleanClause.Occur.MUST);  
Query query = builder.build() ;
```

- Members 의 수가 100 ~1000 인 “서울에 있는 호텔” 질의어로 검색이 되는 도큐먼트를 찾는다.

Index Searcher 를 이용하여 검색 결과 얻기

```
Topdocs tops= searcher.search(query, maxpage);
```

```
for(int i=0 ; tops != null && i<tops.totalHits ; i++) {
```

```
    Document doc = searcher.doc(tops.scoreDocs[i].doc);
```

```
    String id = doc.get("id") ;
```

```
    String name= doc.get("name") ;
```

```
    String city = doc.get("city") ;
```

```
    IndexableField membersF = doc.getField("members") ;
```

```
    Number members = membersF.numericValue() ;
```

```
    int m = members.intValue() ;
```

```
    ::::
```

```
}
```

Search with sort

```
SortField sf = new SortField("pubdate",SortField.STRING,true);
```

```
Sort sort = new Sort(sf);
```

```
hits = searcher.search(query,maxpage,sort);
```

```
SortField sf = new SortField("members",SortField.Type.INT,true);
```

```
Sort sort = new Sort(sf);
```

```
tops = searcher.search(query, maxpage, sort);
```

Field Highlighter

```
QueryScorer scorer = new QueryScorer(query, reader, fieldName);  
Highlighter highlighter = new Highlighter(scorer);  
Fragmenter fragmenter = new SimpleSpanFragmenter(scorer);  
highlighter.setTextFragmenter(fragmenter);
```

```
String fragment = null;  
try {  
    fragment = highlighter.getBestFragment(analyzer, fieldName,  
                                           "서울에 있는 호텔") ;  
} catch (Exception e) {  
    e.printStackTrace();  
    fragment = null;  
}
```


Lucene/Solr 해외 적용 사례

미국 Fortune 지 선정 100 대 기업에서 다수 사용하고 있습니다.

- ❑ Helprace uses Solr 4.0 to power its search engine and search suggestions.
- ❑ Jobreez is powered with Solr 4.0 to search jobs accross 25,000 sources
- ❑ Immonet use Solr 4.0 to real estate search and faceted browsing
- ❑ <http://www.whitehouse.gov/> - Uses Solr via Drupal for site search w/highlighting & faceting (More details: 1)
- ❑ Openindex operates an Apache Solr cluster to provide sitesearch for customers with millions of URL's.
- ❑ Jobuzu search is using Solr search for job search faceted navigation.
- ❑ **Instagram** a **Facebook** company, uses Solr to power it's geo-search API.
- ❑ FCC.gov is the new FCC website featuring Solr powered search and faceted navigation.
- ❑ Comcast / xfinity uses Solr to power site search and faceted navigation.
- ❑ * Jobblu is using as its core search Solr with faceted navigation.
- ❑ **AT&T** Interactive uses Solr to run local search at yp.com, the new yellowpages.com

Lucene/Solr 해외 적용 사례

- ❑ **AOL** is using Solr to power its channels
 - Yellow Pages
 - Music
 - NFL Sports
 - AOL Recipes
 - Real Estate
 - Autos
 - Travel
 - StyleList
- ❑ eHarmony uses Solr for its search features
- ❑ Sears uses Solr for search, auto-suggest, and faceted navigation
- ❑ Ticketmaster is Solr powered, handling search, browsing, alerts, partner feeds.
- ❑ GameSpot uses Solr for its search features.
- ❑ The Guardian uses Solr to power it's Open Platform
- ❑ Netflix uses Solr for their site search feature
- ❑ Chegg uses Solr for their site search feature for services including textbook rental and homework help

Lucene/Solr 해외 적용 사례

- ❑ CNET Reviews and Shopper.com use Solr for product search and faceted browsing
- ❑ Zappos.com uses Solr & SolrJ for their search and navigation across 150,000 styles of shoes and other products.
- ❑ Pink Taffy Designs, a unique baby boutique, uses Solr for search across its entire selection of unique baby gifts.
- ❑ Miinto.com uses Solr for their search and navigation across their network of thousands of brick and mortar stores.
- ❑ SourceForge uses Solr to provide faceted search across all its projects.
- ❑ <http://news.com> uses Solr for search.
- ❑ Panasonic Europe uses Solr to power the search and faceted navigation on its sites for 30 countries.
- ❑ eBay uses Solr to power the search for its German Classified sites.
- ❑ digg uses Solr for search
- ❑ Buy.com's international sites are powered by Solr (LucidWorks), including <http://ca.buy.com> (Canada); <http://fr.buy.com> (France); <http://de.buy.com> (Germany) and <http://uk.buy.com> (United Kingdom).

Lucene/Solr 해외 적용 사례

- ❑ Internet Archive is an opensource digital archive that uses Solr for all of its collections (audio/video/books) search, along with PHP and mysql.
- ❑ StubHub!, an eBay company, uses Solr for Browsing and Searching tickets
- ❑ Smithsonian Institution cross catalog faceted search
- ❑ Homeland Security Digital Library (HSDL) uses Solr for search.
- ❑ NASA Planetary Data System (PDS) uses Solr to search for dataset, mission, instrument, target, and host information
- ❑ Trove service of the National Library of Australia
- ❑ JSTOR uses Solr to power it's search and navigation. JSTOR archives and provides access to archival and current issues of more than 1,400 scholarly journals across more than 50 academic disciplines.
- ❑ GolfVex uses Solr to power site search and to power the online golf score tracker.
- ❑ Evocatus uses Solr for its search features
- ❑ PriceSpin.co.uk UK price comparison website uses Solr for keyword search and faceted search drill-down for the site navigation and search functionality.

Lucene/Solr 해외 적용 사례

- ❑ PriceSpin.co.uk UK price comparison website uses Solr for keyword search and faceted search drill-down for the site navigation and search functionality.
- ❑ Evaluestores uses Solr to power lightning fast search across its massive selection of furniture, home decor, lighting and chandeliers.
- ❑ Ebuyer.com, a UK based computer & electronics retailer uses Solr for its product search
- ❑ FanSnap uses Solr to help fans search for tickets to live sports, concert, and theater events.
- ❑ Car Part Price Comparison - Carmio uses Solr for its complex search. (over 3 mio. products and all car parts are added to a specific car)
- ❑ DealOyster Coupons uses Solr for coupon/merchant search. (3,000+ merchants and 50,000+ coupons)
- ❑ Price Wiki uses Solr for searching cost of living database on their website.
- ❑ <http://www.lucidimagination.com/search> Indexes the Lucene ecosystem: Solr, Lucene Java, Mahout, Nutch, Tika, Droids and the ports of Lucene. Includes mail archives, websites, JIRA issues and wikis.
- ❑ Fashiolista the leading online social network for fashion uses SOLR for tag search, faceted navigation and user search. Uses the Dataimport handler and multicore features.

Lucene/Solr 해외 적용 사례

- ❑ FindBestOpenSourceLists best open source across all categories.
- ❑ <http://www.plaxo.com> - uses Solr for sitewide activity and people search.
- ❑ datingsearcher - uses Solr for searching
- ❑ iOffer is a global marketplace that leverages Solr for its search, faceting and auto-suggest capabilities
- ❑ <http://factbook.bodukai.com/> - CIA World Factbook data imported into Solr.
- ❑ MMtoolparts.com uses Solr for search our huge selection of power tool parts, accessories, hand tools and equipment from all major industrial brands
- ❑ Mefindcoupon uses Solr to help online shoppers search Merchants.
- ❑ Wickedin is a vertical search engine for classified ads launched in Italy and present in Spain and UK. The website successfully uses SOLR to index over 4 million ads in categories such as jobs, real estate, cars, motorbikes, pets, camper vans and general merchandise.
- ❑ Jetwick - an open source twitter search prototype based on Solr and Wicket (featurelist) with a lot of facets. Only one index at the moment hosted from a Virtual Server.
Nevertheless: fast!
- ❑ GittiGidiyor.com is biggest e-commerce company in Turkey and uses Solr for search.

Lucene/Solr 해외 적용 사례

- ❑ HurriyetOto.com - Turkey's leader first & second hand car sale site, build on top of Solr.
- ❑ Kanunum.com is a Turkish enterprise specializing in legal information services and uses Solr for searching Turkish legal texts.
- ❑ All Plumbing Repair has deployed solr to optimize locating plumbers by zip code.
- ❑ Atlanta Water lines Uses Solr for there online shoppers in Atlanta.
- ❑ alaTest.com Product review site (and mobile applications) using Apache Solr for its site search features.
- ❑ The Echo Nest uses Solr quite extensively to store and retrieve billions of data points about music via our API
- ❑ <http://www.alltreatment.com/> Uses Solr to power search and a directory for drug rehab centers nationwide (US).
- ❑ eBharatJobs.com A vertical job search engine targeted for India. Using Solr to index & perform search
- ❑ SimplyFreshers.com A website for needs of Freshers (in India). Using Solr to index & perform search on jobs, placement papers, interviews. Using filters to further refine searches

Lucene/Solr 해외 적용 사례

- ❑ Sarkari Tenders - Indian Government Tenders A vertical tenders search engine targeted for India. Using Solr to index & perform search, and filters results.
- ❑ TheBigJobs.com A Job portal build using Drupal CMS. Using solr to index & search jobs posted on website.
- ❑ Pubget indexes the worlds medical related scientific journals (Medline) together with full-text PDF locations and university library holdings rules. It uses release 1.3 with shards.
- ❑ Anta Plumbing has deployed solr for there company in Toronto.
- ❑ Demand Jobs developed upon Solr which is used to index jobs and provide a fast search interface for users.
- ❑ Citysearch uses Solr for all searches on the site. Includes facets by location and category. Index contains 19 million documents and is distributed over multiple servers. Serving over 3 million visitors a day.
- ❑ IncredibleAgents uses Solr for searching across 1.5 million real estate agents
- ❑ Homethinking uses Solr to search and rank 1.5 million Realtors according to their sales history across each local area

Lucene/Solr 해외 적용 사례

- ❑ manta.com uses Solr for company profiles search and browsing. Facets by location and industry and filtering by sales volume, the number of employee and public/private. Has 14 Million US companies. About 10 requests per second. Running under tomcat 6.x with 6 G JVM Memory (3 Linux servers). The optimized index size is 7.3 G.
- ❑ Game Change change your games for free (Germany)
- ❑ Scissor Doors uses Solr for search and browsing that provides Lambo Doors & Suicide Doors to merchants in USA & Canada.
- ❑ Yankee Group uses Solr to create a search and browse experience which includes its published subscription content, press releases, and public blog postings. Registered users (including guests) can save searches, turn them into email alerts, and consume them as RSS feeds.
- ❑ Central Desktop Multi-tenant SaaS business collaboration tool. >3M documents >1TB raw data >35GB index size.
- ❑ 6pm.com uses Solr for their entire site search and navigation
- ❑ Ridgeway Mechanical Atlanta Plumbing Company. > Solr for their entire site.

Lucene/Solr 해외 적용 사례

- ❑ Enormo - uses Solr to search more than 3,000,000 property listings in 49 European and North African countries, in 30 languages, features range and minimum and maximum percentile faceting
- ❑ Intervals uses Solr for searching projects, tasks and other project management related data.
- ❑ search.com queries various Solr search collections
- ❑ Cleaning Service Directory: Uses Solr to power search over thousands of contractors
- ❑ Atlanta Plumbing uses Solr for there online plumbing site.
- ❑ Volunteer Solutions uses Solr for basic keyword search and faceted browsing of 100s of categories of volunteer opportunities
- ❑ <http://krugle.com> uses Solr to store information about projects and user-created code notes.
- ❑ CHOWHOUND uses Solr for its search facility.
- ❑ Surround Sound DJ uses Solr for speed for there dj site in Toronto.
- ❑ Local Clinical Trials is a clinical trials search engine that uses Solr for trial and facility spatial search.
- ❑ Discogs uses Solr for keyword search and faceted browsing

Lucene/Solr 해외 적용 사례

- ❑ Uptake Vacations: Uses Solr to search over 10,000,000 travel listings of different types including hotels, things to do, reviews, and opinions behind a ruby on rails front end.
- ❑ Collex Collex @ NINES. Collex is a faceted browsing system which Solr behind a Ruby on Rails front-end, with some custom request handlers and caching.
- ❑ ICT Vacatures a Belgian job board for IT and telecommunication jobs powered by Solr.
- ❑ Intellog search, content contribution and sharing site for industrial communities (initially, oil & gas). Utilizes Solr implemented on Amazon Web Services.
- ❑ Tokenizer: Shopping Price Engine - web crawler and search engine.
- ❑ The tsr.ch vidéo player page uses a Solr index as a back-end for rich navigation features. There's some additional info on Bertrand's weblog
- ❑ <http://www.railshostinginfo.com/> uses Solr for its faceted search to find Ruby on Rails hosting.
- ❑ University of Alberta Libraries Peel's Prairie Provinces uses Solr for full text search and faceted search.
- ❑ University Laval library catalogue (1.8 M docs) <http://ariane2.bibl.ulaval.ca>

Lucene/Solr 해외 적용 사례

- ❑ La Repubblica Newspaper - One of the main Italian Newspapers uses Solr for faceted browsing/filtering through classifieds (Annunci)
- ❑ Instructables uses solr for faceted browsing and search. Logged in users get a nice solr-powered library interface.
- ❑ Photosite uses Solr for site search.
- ❑ Zvents uses Solr for search and faceted browsing of events & activities listings.
- ❑ PriceJunkie is a price comparison engine that uses solr to handle search and faceted browsing of merchant products.
- ❑ reddit uses Solr for search
- ❑ UNT Libraries' Digital Collections uses Solr for search features
- ❑ The Portal to Texas History uses Solr for search features
- ❑ The Congressional Research Service Archive uses Solr for search features
- ❑ Scintilla: search and MoreLikeThis (Drupal Solr module)
- ❑ MAME Reviews: faceted search (Drupal Solr module)
- ❑ Peel Sessions: search (Drupal Solr module)
- ❑ Golfballs.com uses Solr for all site browsing functions
- ❑ Howtoons.com uses a solr powered blog.

Lucene/Solr 해외 적용 사례

- ❑ Nsyght uses Solr for search
- ❑ Matratzen-Komet.de uses solr for finding products and furniture for bedroom like beds.
- ❑ Bazaaria is an online trading alternative, focussing on the niche of non-powersellers. Bazaaria uses Solr for its product search and filters.
- ❑ ReviewGist is a product reviews search and comparison tool that uses Solr for faceted browsing and review search.
- ❑ The Digital Commonwealth uses a solr based OAI-PMH engine to search digital collections within Massachusetts.
- ❑ Cosmotourist uses Solr for search and faceted browsing.
- ❑ Flix55 uses Solr to search user-uploaded videos
- ❑ Hitmeister uses Solr for search
- ❑ homestars.com uses Solr to search 1.6 million companies.
- ❑ Hood.de uses Solr for product and auction search.
- ❑ Snooth uses Solr for search.
- ❑ DollarDays uses Solr for product and CRM searches handling tens of thousands of requests per day against over 1.5 million records using Apache-Tomcat.
- ❑ TVtrip use Solr for hotel search and faceted browsing

Lucene/Solr 해외 적용 사례

- ❑ IS Inc. uses Solr for course search and browsing
- ❑ Well.ca uses Solr for product searches
- ❑ Autoo.ro Romanian car sales aggregator. Uses Solr for search and faceted browsing.
- ❑ Rez.ro Romanian real estate portal. Uses Solr for search and faceted browsing.
- ❑ www.sladsonline.com uses Solr to index and search Sri Lankan advertisements published online.
- ❑ The Dewey Browser was developed by the OCLC office of Research and uses Solr to provide resource discovery based on the Dewey Decimal Classification.
- ❑ StyleFeeder: uses Solr for search
- ❑ Avvo: uses Solr for search and faceted browsing.
- ❑ easy-coding.de: uses Solr for onsite and crosssite search, home of solr wcf plugin, ~4k hits a day, uses faceted browsing.
- ❑ Polyvore: uses Solr for faceted search and browsing of items, sets, and groups.
- ❑ situsbelanjaku: uses Solr for faceted search and browsing of items, sets, and groups.
- ❑ Sapo-Videos: uses Solr for content search and to suggest related videos.
- ❑ Sapo-Blogs: uses Solr to search through a collection of 6.5 million documents.

Lucene/Solr 해외 적용 사례

- ❑ Shopreflex use Solr for content search and faceted search
- ❑ Staragora use Solr for content search and faceted search
- ❑ Stilivo uses Solr for faceted browsing for shoes, clothing.
- ❑ <http://dlib.nyu.edu/findingaids/> uses Solr to search among and within archival finding aids for NYU and N-YHS
- ❑ bulDinle.com use Solr for music content search
- ❑ Imoo Romanian real estate listings aggregator. Uses Solr for search and faceted browsing.
- ❑ Wego Travel Search Engine, use Solr to index data for fast retrieval
- ❑ Nordsee-Küchen uses Solr for search in Magento Commerce for finding over 2000 parts like Armaturen, Spülen and Mülleimer
- ❑ Oshieru Manaberu Matchmaking site uses Solr to find teachers and students, each other.
- ❑ Onmeda German health portal uses Solr for content search, including faceted searches. (~750.000 articles, ~9000 searches per day, JBOSS)
- ❑ <http://www.pagenstecher.de> uses Solr do search usergenerated Content, more than 2 Mio documents and 5000 searches a day, using faceted search. Try it out
- ❑ <http://www.rallyformusic.com> uses Solr for navigation and full site search.
- ❑ Fitnessfootwear.com uses Solr for faceted browsing of Hunter, Merrell, Caterpillar. Dual CPU/2GB RAM.
- ❑ <http://www.machdudas.de> uses Solr for location based classified search using geodata.

Lucene/Solr 해외 적용 사례

- ❑ <http://uaprom.net> and <http://ruprom.net> Ukrainian and Russian B2B sites. Use Solr for search and faceted browsing of products, companies and buying leads.
- ❑ <http://www.kerevizli.com> uses Solr to search music albums.
- ❑ Cheaptickets.com Uses solr for faceted search. Searches all the hotel reviews for each location and provides the results to the user.
- ❑ WebCity uses Solr for faceted browsing of restaurants, hotels, events...
- ❑ shoe-shop.com uses Solr and SolrSharp for faceted category browsing on microsites such as <http://www.adidastrainers.co.uk> and <http://www.fashiontrainers.co.uk> to provide search by colour & size.
- ❑ flug.ideal.de uses Solr for faceted browsing of airports, airlines and flight-offers and some searches (Ajax Suggest). Solr (with Jetty) runs on a 8 CPU (2 x E5450 @ 3.00GHz) machine with 32 GB of RAM, 16 are dedicated to Jetty/Solr. Actually (Oct 2008) we have between 200,000 and 500,000 Requests from faceted browsing (up to 15/sec) and 100,000 Searches per day. We have 15,000,000 Docs and faceted browsing is very fast.
- ❑ Firm-UA.com uses Solr for site search
- ❑ CiteSeerX uses Solr as the Index for full text, citation similarity search.
- ❑ YAY Micro: uses Solr to search in over 500 000 images and vector files. Solr has given YAY one of the fastest searches in the stock image industry.

Lucene/Solr 해외 적용 사례

- ❑ podcast.de: German podcasting portal uses Solr for channel, episode, user search and faceted browsing
- ❑ <http://www.airliners.net>: Aviation photography website.
- ❑ findthatfile.com: File based search engine for locating media on the Internet encompassing Web, FTP, Usenet and P2P resources including 43 file types and 200+ file extensions. Last count 300m records. Derivative vertical sites include : <http://www.findthatdoc.com>, <http://www.findthatexcel.com>, <http://www.findthatpdf.com>, <http://www.findthatdll.com>, <http://www.findthatzipfile.com>, <http://www.findthatpowerpoint.com>, <http://www.findthatrpm.com>, <http://www.findthatjar.com>, <http://www.findthatcode.com>, <http://www.findthatdb.com>, <http://www.findthatexe.com>, <http://www.findthataudio.com>, <http://www.findthatvideo.com>
- ❑ buddyfetch: Social network, IM, and dating people search engine (approx 70 million people)
- ❑ Shopr.com uses Solr for keyword and faceted search of its users and user recommended products.
- ❑ :::::
- ❑ more and more

Lucene/Solr 국내 주요기업 적용 사례

커뮤니티 중심에서 대기업 적용으로 늘어나는 추세입니다.

- ❑ KT : ShoDoc : 종합 쇼핑몰 검색 : Solr 분산 검색
- ❑ KTH : DAISY (Data Intelligence System) : 대용량 데이터 실시간 분석·추천 솔루션
- ❑ LG CNS : Smart SMA : 소셜 미디어 분석 솔루션 검색
- ❑ 특허정보원 : 판례등 법률문서 문장 검색

END

Q&A