

SAN FRANCISCO MAY 25-26 2011

Extending Solr: Building a Cloud-like Knowledge Discovery Platform

Trey Grainger, 



LUCENE REVOLUTION



Overview

CareerBuilder's Cloud-like Knowledge Discovery Platform

- **Scalable approaches to multi-lingual text analysis (with research study)**
 - *Multiple fields vs Multiple Cores vs Single Field*
- **Custom Scoring**
 - *Payloads and on-the-fly bucket scoring*
 - *Implementing a keyword spamming penalty*
- **Solr as a Cloud Service**
 - *Scalable, customizable search for everybody*
- **Knowledge Discovery & Data Analytics**

My background

- **Trey Grainger**
 - **Search Technology Development Team Lead @ CareerBuilder.com**
- **Relevant Background:**
 - **Search & Recommendations**
 - **High-volume, N-tier Architectures**
 - **NLP, Relevancy Tuning, user group testing & machine learning**
- **Fun Side Project:**
 - **Founder and Site Architect @ Celiaccess.com**

CareerBuilder's Search Scale

- Over 1 million new jobs each month
- Over 40 million resumes
- ~150 globally distributed search servers
(in the U.S., Europe, & Asia)
- Several thousand unique, dynamically generated indexes
- Over a million searches an hour
- >100 Million Search Documents

Job Search



cbsalary
Find Salary Information
for these Jobs >>

Save this search
View saved jobs
Tell CareerBuilder!

Lucene jobs

1 - 25 of 67 jobs found

Search Again Advanced Search

Keywords: Lucene Location: e.g. "Chicago, IL" or "60601" Posted within: Last 30 Days [Find Jobs »](#)

Narrow Search

Category:	Company:	City:	State:
Information Technology (62)	CyberCoders Engineering (16)	New York (10)	California (20)
Engineering (44)	IBM USA (8)	McLean (4)	New York (11)
Professional Services (18)	Dex One (5)	Durham (3)	Virginia (8)
Design (14)	Technisource (5)	Glendale (3)	Illinois (5)
Other (4)	Bronto Software (3)	Brisbane (2)	North Carolina (5)

Show me: [Closest jobs first](#) | [Relevant jobs first](#) | [Newest jobs first](#)

Lucene Search Results

Page 1 of 3 | [Next Page](#)

Job Title / Description (show titles only)	Company	Location	Posted
Software Engineer - .Net/Java - Lucene - Solr - Compass - View similar jobs Job type: Full-Time This position is open as of 5/16/2011. Software Engineer - .Net/Java - Lucene - TFS - Droids ... Lucene, TFS, .Net, J2EE, Java, ... View full job description Save to MyCareerBuilder Email to a friend	CyberCoders Engineering	TX - Kingwood	4 Days Ago
GIS J2EE Software Developer - J2EE/JAVA Lucene Developer - View similar jobs Job type: Full-Time This position is open as of 5/16/2011. J2EE Lucene Developer - JAVA / J2EE GEO / GIS ... JAVA, J2EE, Apache Lucene, TFIDF, Google Maps... View full job description Save to MyCareerBuilder Email to a friend	CyberCoders Engineering	CA - El Segundo	4 Days Ago
Java Developer: Lucene - View similar jobs Job type: Contractor developer position requires an understanding of Lucene, and experience working with large datasets. The... View full job description Save to MyCareerBuilder Email to a friend	Manpower Professional	CA - Palo Alto	1 Week Ago

Resume Search



[Back to Search Results](#)

ResumeFlip^{beta}

YOUR SEARCH

lucene in All edited within the last 90 days

2 of 25

[Feedback](#)

Home My Career

Jobs & Application

Current Search:

[Feedback](#) | [Save Search](#)

More Search

Keywords

lucene

Using

All of these

Locations

30 Miles

e.g. Chicago, IL

Add Another

Include only relevant

Maximum commute

0

Last Modified

Within 3 months

Company Name

[Redacted Name]
[Redacted Email]@[Redacted].edu

[Redacted Phone]

Cell: [Redacted]

SUMMARY

Seeking for a *Software Engineer* position.
With 1 year research experience in *Apache Lucene search engine* and 1 year Graduate Assistant in *National Instrument (NI) machine controlling in Visual Basic*
Diverse knowledge in *SQL query, indexes, database optimization, data structure, and automata theory*

EDUCATION

[Redacted] University, [Redacted], [Redacted]
Master of Science in Computer Science 2009-2011
[Redacted] University, [Redacted], [Redacted]
Bachelor of Science in Computer Science and Information Engineering 2002-2006

PROFESSIONAL EXPERIENCE

- [Redacted] University, [Redacted], [Redacted]*
Research Assistant, Chemistry and Biochemistry Department 2009-2010
- Researched and developed the National Instrument (NI) machine functionality including Visual Basic program for PCI-card voltage detection, and automation control technique
 - Modified Labview figure processes, implemented Visual Basic program to control the Labview functionality, and maintained instruments

Block

Print

[View Full Profile](#)

new resume viewer!

Page: 1 2 ... 7

View By:

[Similar](#) [Block](#)

Talent Network Search



careerbuilder[®]
For Employers

Hello, Trey Grainger (not you?) [Site Security](#) [Help](#) [Sign Out](#)

Learn about best practices
for ensuring top quality hires.

[Click Here.](#)

Personified
a Division of Career Builder

[Home](#) [My CareerBuilder](#) [Products](#) [Solutions](#) [Resources](#) [\\$0 Checkout](#) [Release RDB License](#)

[Jobs & Applications](#)

[Resume Search](#)

[Talent Intelligence Portal](#)

[Account Tools](#)

[Jobs & Applications](#) > [Talent Search](#) > [Talent Search Results](#)

- [Post a Job](#)
- [View My Jobs](#)
- [Screeners](#)
- [Applications](#)
- [Talent Search](#)
- [My Templates](#)

More Search Options

Keywords

Locations

Last Modified

Page: 1 2 3 4 ... 50 Next >

1-25 of 1,250

Person Name		Location	Last transaction date
<input type="checkbox"/> [Redacted]	NewJobTitleNO Date Applied: 5/18/2011 9:14:01 AM	Stone Mountain-GA-US	5/18/2011 9:14:01 AM
<input type="checkbox"/> [Redacted]	NewJobTitleNO Date Applied: 5/17/2011 11:12:55 AM	Lithonia-GA-US	5/17/2011 11:12:55 AM
<input type="checkbox"/> [Redacted]	NewJobTitleNO Date Applied: 5/11/2011 8:34:21 AM	Marietta-GA-US	5/11/2011 8:34:21 AM
<input type="checkbox"/> [Redacted]	Test Job by (Not manager) Date Applied: 5/10/2011 3:35:54 PM	Duluth-GA-US	5/10/2011 3:35:54 PM
<input type="checkbox"/> [Redacted]	Test Job by (Not manager) Date Applied: 5/10/2011 12:28:48 PM	Morrow-GA-US	5/10/2011 12:28:48 PM
<input type="checkbox"/> [Redacted]	Test Job by (Not manager) Date Applied: 5/9/2011 7:23:56 PM	Canton-GA-US	5/9/2011 7:23:56 PM
<input type="checkbox"/> [Redacted]	Test Job by (Not manager)		

Auto-Complete



Search Again | Advanced Search

Keywords:
software engineer
software
software sales
software developer
sr software engineer
senior software engineer
software test engineer
software engineering
software test
software manager

Location:

Posted within: Last 30 Days

City: New York (10), McLean (4), Durham (3), Glendale (3), Brisbane (2)

State: California (20), New York (11), Virginia (8), Illinois (5), North Carolina (5)

Show me: [Closest jobs first](#) | [Relevant jobs first](#) | [Newest jobs first](#)

Lucene Search Results

Job Title / Description (show titles only)	Company	
Software Engineer - .Net/Java - Lucene - Solr - Compass - View similar jobs Job type: Full-Time This position is open as of 5/16/2011. Software Engineer - .Net/Java - Lucene - TFS - Droids ... Lucene, TFS, .Net, J2EE, Java, ... View full job description Save to MyCareerBuilder Email to a friend	CyberCoders Engineering	TD
GIS J2EE Software Developer - J2EE/JAVA Lucene Developer - View similar jobs Job type: Full-Time This position is open as of 5/16/2011. J2EE Lucene Developer - JAVA / J2EE GEO / GIS ... JAVA, J2EE, Apache Lucene, TFIDF, Google Maps... View full job description Save to MyCareerBuilder Email to a friend	CyberCoders Engineering	CA
Java Developer: Lucene - View similar jobs Job type: Contractor developer position requires an understanding of Lucene, and experience working with large datasets. The... View full job description Save to MyCareerBuilder Email to a friend	Manpower Professional	CA

Geo-spatial Search

<< Back to Job Results View

 Tell Us What You Think!

 MapBasedSearch BETA

**Note: Only jobs where employers provided addresses are displayed in the map view.

List View Map View

Results 1 - 25 of 931 [Next Page](#)

Keywords: Location:

e.g. Nursing, Banking, etc. Chicago, IL or 60601

1. Senior Software Engineer
Full-Time Employee
Silicon Valley, CA | Miles
[View Description](#)

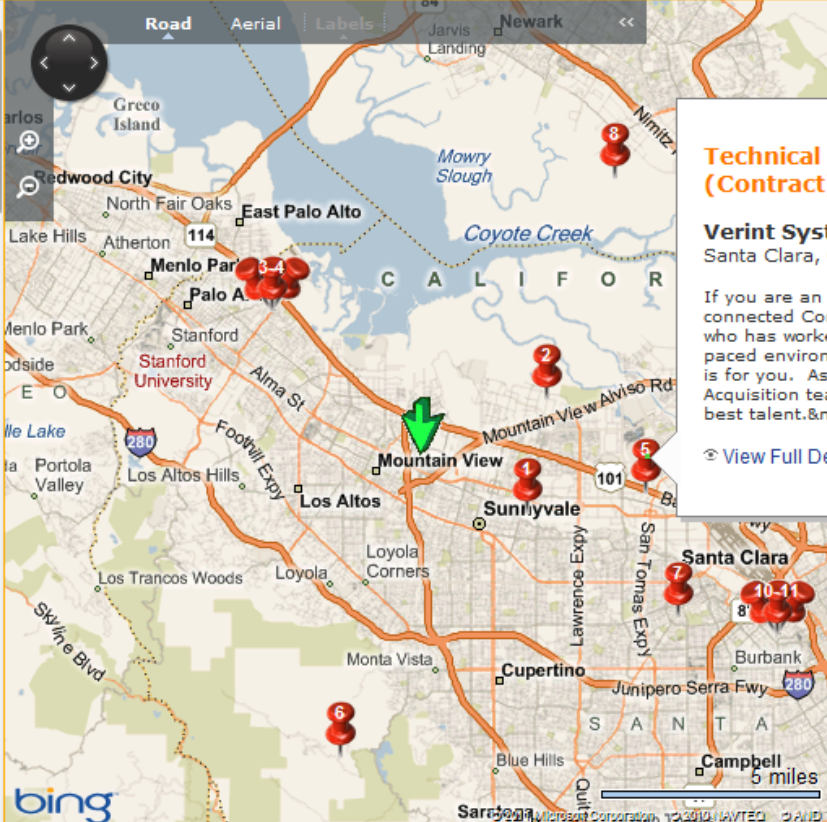
2. Sr. Business Systems Analyst
Full-Time Employee
Sunnyvale, CA | Miles
[View Description](#)

3. Senior Software Engineer (Web Channel Application)
Full-Time Employee
Palo Alto, CA | Miles
[View Description](#)

4. Senior Software Engineer (Platform and services)
Full-Time Employee
Palo Alto, CA | Miles
[View Description](#)

5. Technical Recruiter (Contract)
Part-Time Employee
Santa Clara, CA | Miles
[View Description](#)

6. Software Engineer
Full-Time Employee



Technical Recruiter (Contract)
Verint Systems
Santa Clara, CA | Miles

If you are an experienced and connected Contract Technical Recruiter who has worked in a growing, fast paced environment, then this contract is for you. As a part of our Talent Acquisition team, you will bring in the best talent.&n . . .

[View Full Description](#)

Recommendations

- We classify all content (Jobs, Resumes, etc.) and index the classified content into Solr
- We use a combination of collaborative filtering and classification techniques
- We utilize a custom scorer and payloads to apply higher bucket weights to more relevant content
- Recommendations are real-time and largely driven by search

Job Recommendations



Jobs like Software Engineer - Search Technology Development at CareerBuilder-US

Jobs Like Software Engineer - Search ...	Software Java	Software Microsoft	Software Web Frontend	Data Database	Software Business Analysis	100 Results Found	
Job Title	Company	Number of Applicants	Location	Level	Fit	Date	
Software Engineer NEW!	Verint Systems		Roswell, GA	Mid	★★★★☆	5/18/2011	Similar
Java Developer	Synergis		Atlanta, GA	Mid	★★★★☆	5/11/2011	Similar
Sr. Software Engineer: (Java), GWT, Spring, Hibernate, Oracle NEW!	McKesson		Alpharetta, GA	Mid	★★★★☆	5/17/2011	Similar
Java Developer NEW!	DISYS		Atlanta, GA	Mid	★★★★☆	5/17/2011	Similar
Java Application Developer	Adecco Technical		Alpharetta, GA	Mid	★★★★☆	5/11/2011	Similar
Lead Java SOA Developer NEW!	Volt		Atlanta, GA	Mid	★★★★☆	5/19/2011	Similar
C, C++, Java, PowerBuilder Developers NEW! Atlanta, GA Alpharetta, GA	SetFocus		2 Locations	Mid	★★★★☆	5/16/2011	Similar
Lead Software Engineer/Developer NEW!	Equifax		Alpharetta, GA	Mid	★★★★☆	5/16/2011	Similar
Webmethods Administrator	Pyramid, Inc		Alpharetta, GA	Mid	★★★★☆	5/10/2011	Similar
Software Engineer	NCR		Duluth, GA	Mid	★★★★☆	5/2/2011	Similar
Sr J2EE Developer - Server Side NEW!	COMSYS		Alpharetta, GA	Mid	★★★★☆	5/18/2011	Similar
Java Can Change Healthcare! See how @CareLogistics NEW!	Jackson Healthcare		Alpharetta, GA	Mid	★★★★☆	5/16/2011	Similar
Senior Java / J2EE Developer	<Private>		Alpharetta, GA	Mid	★★★★☆	5/10/2011	Similar
Senior WPF Developer -Silverlight -.NET 4.0 - PAID RELOCATION NEW!	CyberCoders Engineering		Atlanta, GA	Mid	★★★★☆	5/16/2011	Similar
C# .NET DEVELOPER - IMMEDIATE NEED! NEW!	Technisource		Atlanta, GA	Mid	★★★★☆	5/17/2011	Similar

Resume Recommendations



[Resume Search](#) > [Resume Match](#) > Recommended Resumes

Resume Match

[New Search](#)

100 resumes found that match the resume **Large-scale distributed systems engineer**

Your Search Criteria [Modify](#) Customize and rank Areas, Job Titles, & Skills to get better matching results

<p>Areas</p> <p>Java (38), Artificial Intelligence (1), Web Backend (24), Business Analysis (72), Web Frontend (24), Unix (25), Database (27)</p>	<p>Job Titles</p> <p>Systems Engineer (28)</p>	<p>Skills</p> <p>Java (35), Systems (68), Machine Learn (0), Jdbc (9), Tomcat (11), Jsf (4), Junit (11), Jsp (16), Eclipse (16), Web Service (9)</p>
--	---	---

Refine Your Results

Location

Within: miles of

Last Modified

Within: days

- [Job Duration](#)
- [Education Level](#)
- [Experience Level](#)
- [Activity Level](#)
- [Management Level](#)

[Update Your Results](#)

Results

Action Sort By: [Load Saved Search](#) [Save this Search](#)

- Java/J2EE_Developer** [View Similar](#)

_____, Bailey CO Last Modified 5/2/2011

<input type="checkbox"/> RELEVANCE	★★★★★	RECENT JOB	Sr. Java Developer	EXP. LEVEL	Senior
<input type="checkbox"/> ACTIVITY	Inactive	EDUCATION		JOB DURATION	Short
				MGMT EXP	None
				DEGREE TYPE	4 Year

AREAS	JOB TITLES	SKILLS
Java Database Web Backend	Java Developer, J2ee Developer, Software Consultant	Java, J2ee, Junit, Eclipse, Jsp, Tomcat, Ejb, Jsf, Bea, Weblogic,

[View Similar](#)

Last Modified 5/2/2011
- WeblogicAdmin** [View Similar](#)

_____, Longmont CO Last Modified 5/16/2011

<input type="checkbox"/> RELEVANCE	★★★★★	RECENT JOB	WebLogic Administrator / Support	EXP. LEVEL	Mid
<input type="checkbox"/> ACTIVITY	none	EDUCATION		JOB DURATION	Long
				MGMT EXP	None
				DEGREE TYPE	Hs

AREAS	JOB TITLES	SKILLS
Java Unix Database	Weblogic Administrator, Tomcat Administrator	Weblogic, Bea, Java, J2ee, Jms, Ejb, Jsp, Jdbc, Servlet, Junit, Jsf,

[View Similar](#)

Last Modified 5/16/2011
- _resume** [View Similar](#)

_____, Greenwood Village CO Last Modified 5/22/2011

<input type="checkbox"/> RELEVANCE	★★★★★	RECENT JOB	J2EE Developer	EXP. LEVEL	Mid
------------------------------------	-------	------------	----------------	------------	-----

[View Similar](#)

Last Modified 5/22/2011

Resume Icons

- You have previously viewed this resume.
- This resume has been saved to one of your folders.

Multi-lingual Analysis

■ Approach 1: Different Field Per Language

- **Advantages:**

- *Simple, easiest to implement*

- **Disadvantages:**

- *My require keeping duplicate copies of your text per language*
- *If searching across each field (dismax style), slows search down, especially if handling many languages*

■ Approach 2: Different Solr Core per language

Each core has your field defined with a different Analyzer chain specific to that core's language

- **Advantages:**

- *Searching can be completely language-agnostic and additional overhead to search more languages simultaneously is negligible*

- **Disadvantages:**

- *Multi-lingual documents require indexing to multiple cores, potentially messing up relevancy and adding complexity*
- *Have to write your own language-dependent sharding*
- *If you don't already have distributed search, this adds complexity and overhead*

Multi-lingual Analysis

- Approach 3: All languages in one field
 - **Advantages:**
 - Only one field needed regardless of number of languages
 - Avoids a field explosion or a Solr core explosion as you scale to handle more languages
 - **Disadvantages:**
 - Can end up with some “noise” in the index if you process most text in lots of languages (especially if stemming and not lemmatizing)
 - Currently requires writing your own Tokenizer or Filter
- Strategy:
 - 1) Copy token stream and create a stemmer/lemmatizer for each language
 - 2) Pass the original into each stemmer/lemmatizer
 - 3) Stack the outputs of each stemmer/lemmatizer

Input:	en,es,el Εάν αναζητάτε μια λαμπρή επαγγελματική καριέρα, CareerBuilder is one of the US compańias hiring search engineers																
Output:	εάν	αναζητάτε	μια	λαμπρή	επαγγελματική	καριέρα	career	builder	is	one	of	the	us	compańias	hiring	search	engineers
	αν	αναζητώ	ένας	λαμπρός	επαγγελματικός			careerbuilder	be					compańia			engineer

Multi-lingual Analysis

■ Case Study: Stemming vs. Lemmatization

- Example: dries >> dri vs dries >> dry

	English		Greek		French		German		Spanish		Italian		Dutch	
	%+	% -	%+	% -	%+	% -	%+	% -	%+	% -	%+	% -	%+	% -
Commercial Search Vendor	+0	-0	+0	-0	+0	-0	+0	-0	+0	-0	+0	-0	+0	-0
Solr (Lemmas)	+4	-3	+84	-6	+12	-5	+18	-8	+9	-5	+5	-13	+15	-2
Solr (Snowball Stemming)	+28	-2	+10	-13	+25	-2	+24	-7	+20	-2	+25	-.5	+17	-4
Nothing	+1	-48	+5	-30	+1	-23	+2	-35	+1	-32	+0	-27	+1	-16
K Stemmer	+7	-11	Measuring Recall Overlap Between Options											

Take-away: Lemmatization allows you to greatly increase recall while preserving the precision you lose with stemming

i.e. English shows 92% increase in recall using Lemmatization with minimal impact on precision

Custom Scoring

- Search Terms can be boosted differently:
 - `q=web^2 development^5 AND jobtitle:(software engineer)^10`
- Some Fields can be weighted (scored) higher than others
 - i.e. **Field1¹⁰, Field2⁵, Field3², Field 4^{.01}**
- Content within Fields can be boosted differently
 - `design [1] / engineer [1] / really [] / great [] / job [] / ten[3] / years[3] / experience[3] / careerbuilder [2] / design [2], ...`
Field1: `bucket=[1] boost=10`; **Field2:** `bucket=[2] boost=1.5`; **Field3:** `bucket=[] weight=1`; **Field4:** `bucket=3 weight=1.5`
 - We can pass in a parameter to solr at query time specifying the boost to apply to each bucket i.e. `...&bucketWeights=1:10;2:1.5;3:1.5`
- You can also do index-time boosting, but this reduces your ability to do query-side relevancy experiments and requires norms to always be on
- By making all scoring parameters overridable at query time, we are able to do A / B testing to consistently improve our relevancy model

Stopping Keyword Spamming

- We already subclass PayloadTermQuery and tie in custom scoring for our buckets weights
- For each payload “bucket” (or across all buckets), we can count the number of hits and penalize the score if a particular keyword appears too many times
- Payload scoring then essentially becomes
 - **BucketBoost(payloadBucket) * HitMap(#hitsPerbucket)**
- By adjusting our HitMap function, we can thus generate any kind of relevancy curve for how much each additional term adds to (or subtracts from) the relevancy score for that document
 - **ex: Bell curve, Linear, Bi-linear, Linear with drop-off, custom map, etc.**

CareerBuilder's Search Cloud



- Goals:
 - Make search easy to use and accessible to all engineers (not just the search team)
 - Allow schema changes without mucking with solr (on hundreds of servers)
 - Make solr installs generic and independent of any particular implementation

Creating a virtual search engine

- 3 Main Cloud Actions: Index, Search, Delete

```
public static void Main() {
    String poolName = "TestPool"; //equivalent of developer key

    //Indexing
    ////1) We must create a Schema object for the Document
    Schema testSchema = CreateTestSchema(poolName);

    ////2) We must create the Document (passing in the Schema)
    Document testDocument = CreateTestDocument(testSchema);

    ////3) We must call AddDocument on the SearchEngine (passing in the Document)
    Boolean addDocumentSuccess = SearchEngine.AddDocument(testDocument);

    //Searching
    SearchCriteria searchCriteria = CreateTestSearchCriteria(testSchema);
    QueryResult queryresult = SearchEngine.Search(searchCriteria);

    //Cleanup
    Boolean deleteSuccess = SearchEngine.DeleteDocument("TestPool", testDocument.DocumentID);
}
```

Creating a virtual search engine

■ Creating a Schema

```
public static Schema CreateTestSchema(String poolName) {
    Schema schema = new Schema(poolName, "TestSchema", SupportedLanguages.English);
    //Let's Create a few Fields to Add to the Schema
    StringField nameField = new StringField("Name");
    nameField.MakeSearchable();
    nameField.MakeRetrievable();
    nameField.MakeSortable();
    nameField.MakeFacetable();
    schema.AddField(nameField);

    DelimitedTextField semiColonDelimitedField = new DelimitedTextField("SemiColonDelimited",
                                                                    DelimitedTextField.Delimiters.SemiColon);

    semiColonDelimitedField.MakeSearchable();
    schema.AddField(semiColonDelimitedField);

    FreeTextField germanEnglishText = new FreeTextField("GermanEnglishText", new SupportedLanguages[]
                                                                    {SupportedLanguages.German, SupportedLanguages.English});

    germanEnglishText.MakeSearchable();
    germanEnglishText.SetBoost(1.2);
    schema.AddField(germanEnglishText);

    BooleanField isHidden = new BooleanField("IsHidden");
    isHidden.MakeSearchable();
    schema.AddField(isHidden);

    return schema;
}
```

Creating a virtual search engine

■ Creating a Document

```
public static Document CreateTestDocument(Schema schema){  
    String randDocId = "T" + (new Random()).Next();  
    Document document = new Document(schema, randDocId);  
    document.GetField("Name").FieldValue = "Trey Test Spaces";  
    document.GetField("SemiColonDelimited").FieldValue = "only;split on;semicolons";  
    document.GetField("GermanEnglishText").FieldValue = "German + English text here";  
    document.GetField("IsHidden").FieldValue = false;  
  
    return document;  
}
```

■ Processing Results

- A **QueryResult** object comes back from the `SearchEngine.Search` method with all of the main types (search records, facets, meta info, etc) parsed out into objects

■ Behind the Scenes:

- We have a distributed architecture handling queuing all documents to appropriate datacenters, feeding the clusters, and load-balancing searches between all available clusters for the given search pool.

Knowledge Discovery & Data Analytics



Home My CareerBuilder Products Solutions Resources \$0 Checkout Release RDB License

Jobs & Applications Resume Search **Talent Intelligence Portal** Account Tools

Talent Intelligence Portal > Supply & Demand Report > Supply Demand Report Summary

Create a New Report

Title, Keyword, or Skill:

software engineer

Location:

California

Timeframe: Last 6 Months

Search

Results For:

Software Engineer in California

[create a new report](#) | [provide your feedback](#)

Data Summary for the Last 6 Months

80,579 Total Unique Job Postings [? What is this?](#)

42,931 Active Resumes [? What is this?](#)

[View These Resumes](#) [? What is this?](#)

There are **0.53** active job seekers for every job posted during the specified timeframe. [? What is this?](#)

Total Available Workforce as of 03/31/2011 [? What is this?](#)

84,996 Computer Software Engineers, Applications

74,427 Computer Software Engineers, Systems Software

38,140 Network And Computer Systems Administrators

34,322 Computer Specialists, All Other

19,093 Industrial Engineers

Filter By

Occupation [?](#)

All Occupations

Computer Software Engineers, Applications (26,521)

Software Quality Assurance Engineers And Testers (5,853)

Computer Software Engineers, Systems Software (5,840)

Web Developers (5,462)

Network And Computer Systems Administrators (4,252)

Show more filters...

Company Name

All Company Names

Cybercoders (4,335)

Cybercoders Engineering (1,825)

Hewlett-Packard Company (1,091)

Adecco S.A. (723)

Google Inc. (579)

Show more filters...

Demand
Job Postings

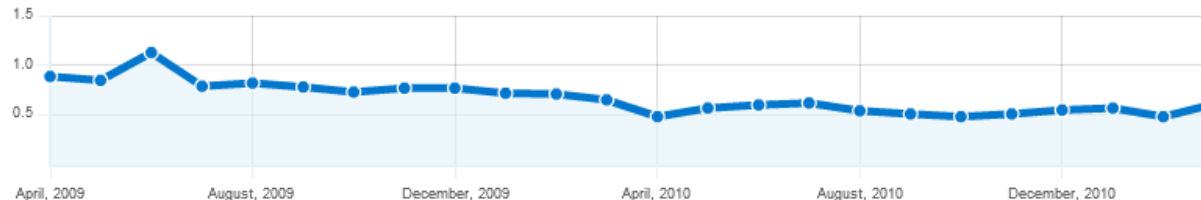
Supply
Workforce - Active and Total

Labor Pressure
Supply Over Demand

Labor Pressure Based on the Active Workforce [? What is this?](#)

Trend Over the Last 24 Months

[View the Full Report](#)



Labor Pressure based on Active Workforce by State

A lower number means more difficult to fill a position and a higher number means less difficult to fill a position. Hover over a state to see exact total

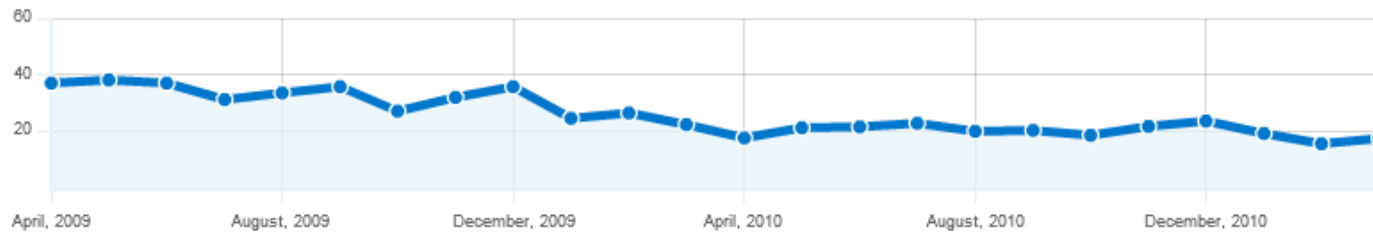
Knowledge Discovery & Data Analytics

Labor Pressure Based on the Total Workforce [? What is this?](#)

Trend Over the Last 24 Months

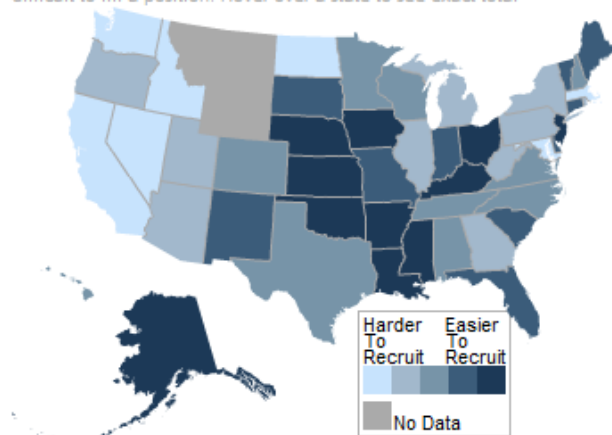
Computer Software Engineers, Applications

[View the Full Report](#)



Labor Pressure based on Total Workforce by State As of 03/31/2011

A lower number means more difficult to fill a position and a higher number means less difficult to fill a position. Hover over a state to see exact total



[View the Full Report](#)

Knowledge Discovery & Data Analytics



Industry

- All Industries
- Temporary Help Services (3,306)
- Computer Systems Design Services (3,258)
- Software Publishers (2,956)
- Semiconductor And Related Device Manufacturing (1,991)
- Custom Computer Programming Services (1,925)

[Show more filters...](#)

Include Staffing

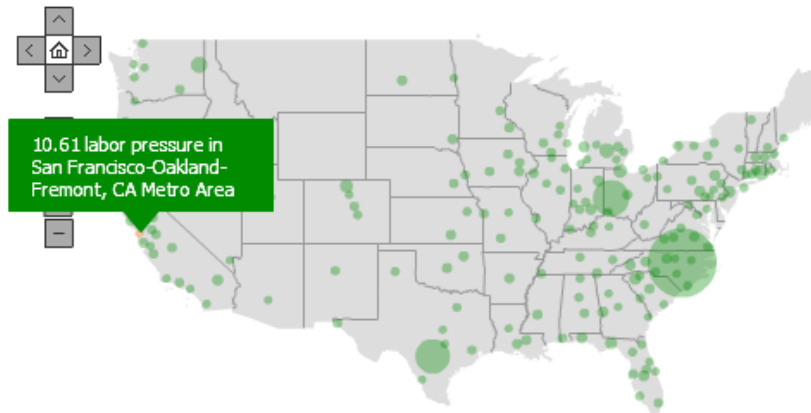
- All (Staffing and Non-Staffing)
- No (60,270)
- Yes (20,309)

Labor Pressure Geography Details

State MSA

Choose a month and year filter to view its historical data.

04/2009	05/2009	06/2009	07/2009	08/2009	09/2009	10/2009	11/2009	12/2009	01/2010	02/2010	03/2010
04/2010	05/2010	06/2010	07/2010	08/2010	09/2010	10/2010	11/2010	12/2010	01/2011	02/2011	03/2011



Hover over a msa to see exact total or clicking on a msa to see its % change

Rank	MSA Name	02/2011	03/2011	% Change
1	Raleigh-Cary, NC	930.00	4,657.00	+400.75 %
2	Dayton, OH	708.67	2,115.00	+198.45 %
3	San Antonio, TX	2,014.00	2,013.00	-0.05 %
4	Spokane, WA	192.00	578.00	+201.04 %
5	Norwich-New London, CT	0.00	545.00	0.00 %
6	Lansing-East Lansing, MI	53.65	465.00	+766.73 %
7	Toledo, OH	83.60	418.00	+400.00 %
8	Fort Collins-Loveland, CO	173.25	346.50	+100.00 %
9	Trenton-Ewing, NJ	427.56	275.14	-35.65 %
10	Tallahassee, FL	87.67	263.00	+199.99 %

Knowledge Discovery & Data Analytics



Create a New Report

Title, Keyword, or Skill:

software engineer

Location:

California

Timeframe: Last 6 Months

Search

Filter By

Occupation

- All Occupations
- Computer Software Engineers, Applications (26,521)
- Software Quality Assurance Engineers And Testers (5,853)
- Computer Software Engineers, Systems Software (5,840)
- Web Developers (5,462)
- Network And Computer Systems Administrators (4,252)

Show more filters...

Company Name

- All Company Names
- Cybercoders (4,335)
- Cybercoders Engineering (1,825)
- Hewlett-Packard Company (1,091)

Results For:

Software Engineer in California

[create a new report](#) | [provide your feedback](#)

Data Summary for the Last 6 Months

80,579 Total Unique Job Postings

Total Available Workforce as of 03/31/2011

84,996 Computer Software Engineers, Applications

Active Labor Pressure Trends

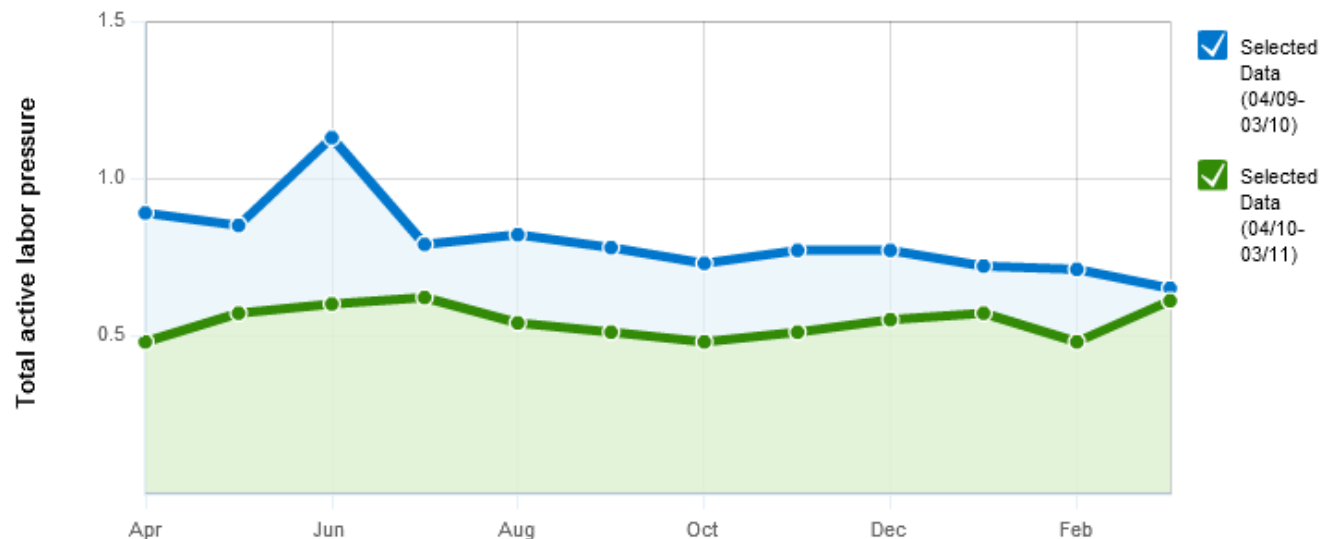
Trend View

Year-Over-Year View

Table View

Show data in percentage change

National data is not shown in absolute terms because its scale is much larger than the selected data.



Knowledge Discovery & Data Analytics

- REPORT DETAILS
- ▶ Methodology
 - ▶ **Data Summary**
 - ▶ Custom View
 - ▶ Related Titles


Registered Nurse

[generate new report](#) [view searched reports](#)

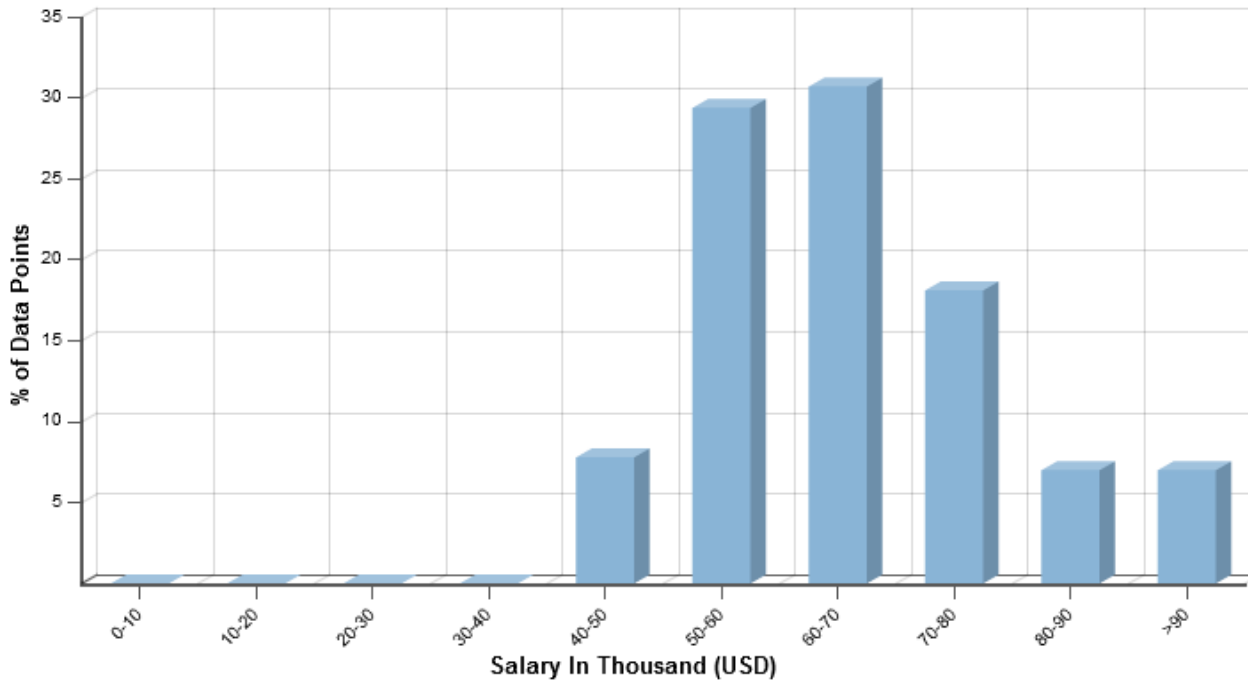
 Print  Export PDF

Salary View Hourly View

Overall Summary

	# Reporting	Avg.	Salary Distribution (percentile) 				
			10th	25th	50th ★	75th	90th
National Average	1,000	\$65,740	\$50,521	\$55,120	\$62,400	\$72,800	\$83,200

Salary Distribution



Knowledge Discovery & Data Analytics



▼ Salary by metropolitan statistical area (msa) ** Figures may not add up to total due to missing or unavailable values*

Metropolitan Statistical Area (MSA)	# Reporting	Avg.	Salary Distribution (percentile) ⓘ				
			10th	25th	50th ★	75th	90th
New York-Northern New Jersey-Long Island, NY-NJ-PA Metro Area	28.64 %	\$77,501	\$60,000	\$65,000	\$75,000	\$88,000	\$101,600
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	16.08 %	\$73,625	\$62,040	\$62,400	\$72,800	\$79,155	\$83,200
Chicago-Naperville-Joliet, IL-IN-WI Metro Area	16.08 %	\$67,970	\$55,000	\$58,055	\$64,303	\$72,200	\$81,912
Dallas-Fort Worth-Arlington, TX Metro Area	14.07 %	\$69,463	\$55,632	\$60,000	\$69,000	\$72,850	\$77,572
Detroit-Warren-Livonia, MI Metro Area	12.56 %	\$61,126	\$51,006	\$54,891	\$60,320	\$66,280	\$71,200
Phoenix-Mesa-Scottsdale, AZ Metro Area	12.56 %	\$67,568	\$56,168	\$60,320	\$66,560	\$72,800	\$79,200

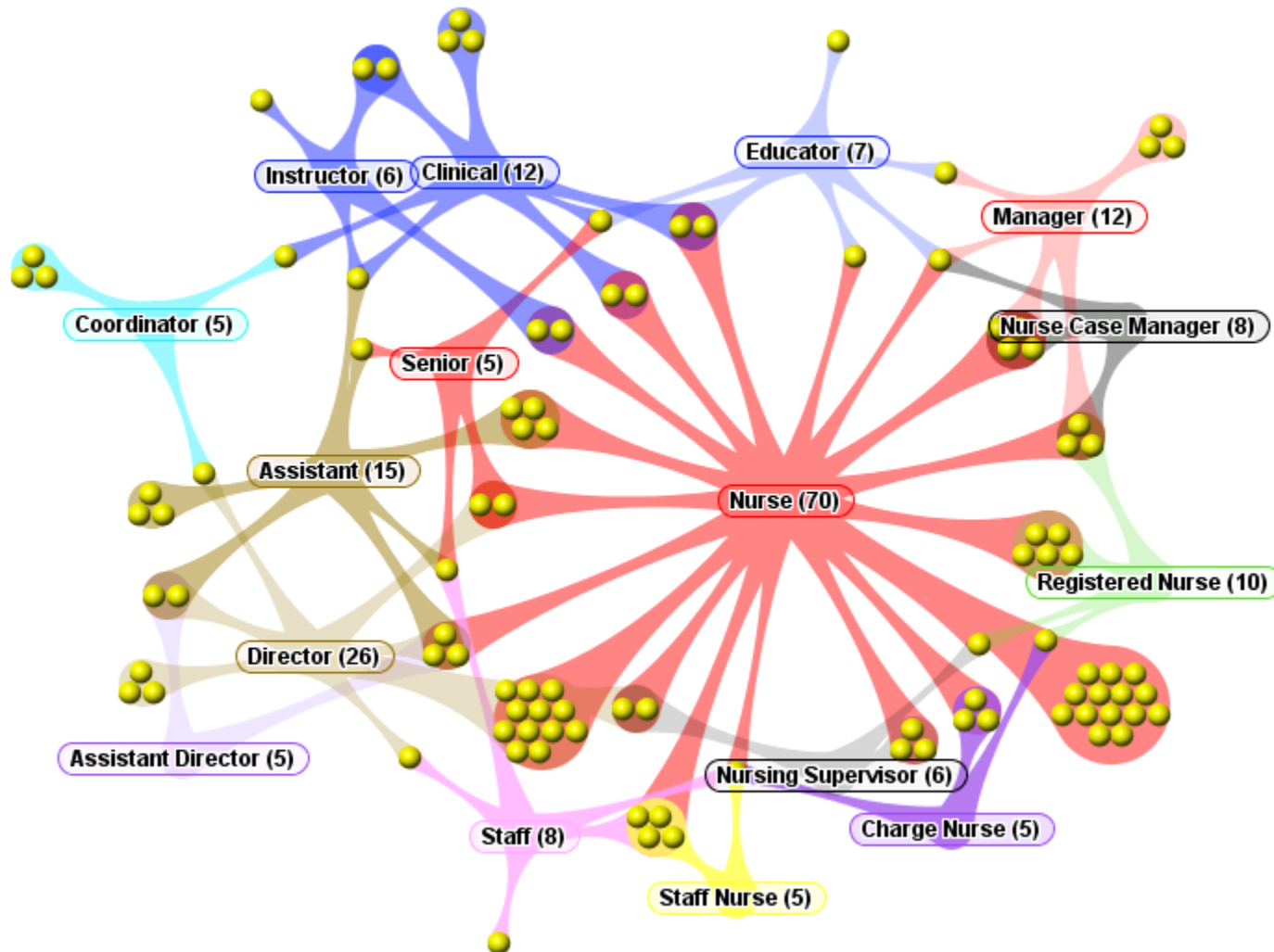
▼ Salary by experience ** Figures may not add up to total due to missing or unavailable values*

Experience	# Reporting	Avg.	Salary Distribution (percentile) ⓘ				
			10th	25th	50th ★	75th	90th
0-1 years	9.58 %	\$62,307	\$49,254	\$51,480	\$60,000	\$71,344	\$80,896
2-5 years	25.06 %	\$64,316	\$50,000	\$54,080	\$62,000	\$72,800	\$85,000
6-10 years	16.63 %	\$63,519	\$50,157	\$54,346	\$60,320	\$69,056	\$79,028
11-15 years	13.28 %	\$64,564	\$52,000	\$55,000	\$62,400	\$70,125	\$79,024
16-20 years	12.70 %	\$66,037	\$52,000	\$58,240	\$62,400	\$71,810	\$81,220
21+ years	22.75 %	\$67,129	\$52,000	\$58,000	\$64,480	\$73,000	\$83,080

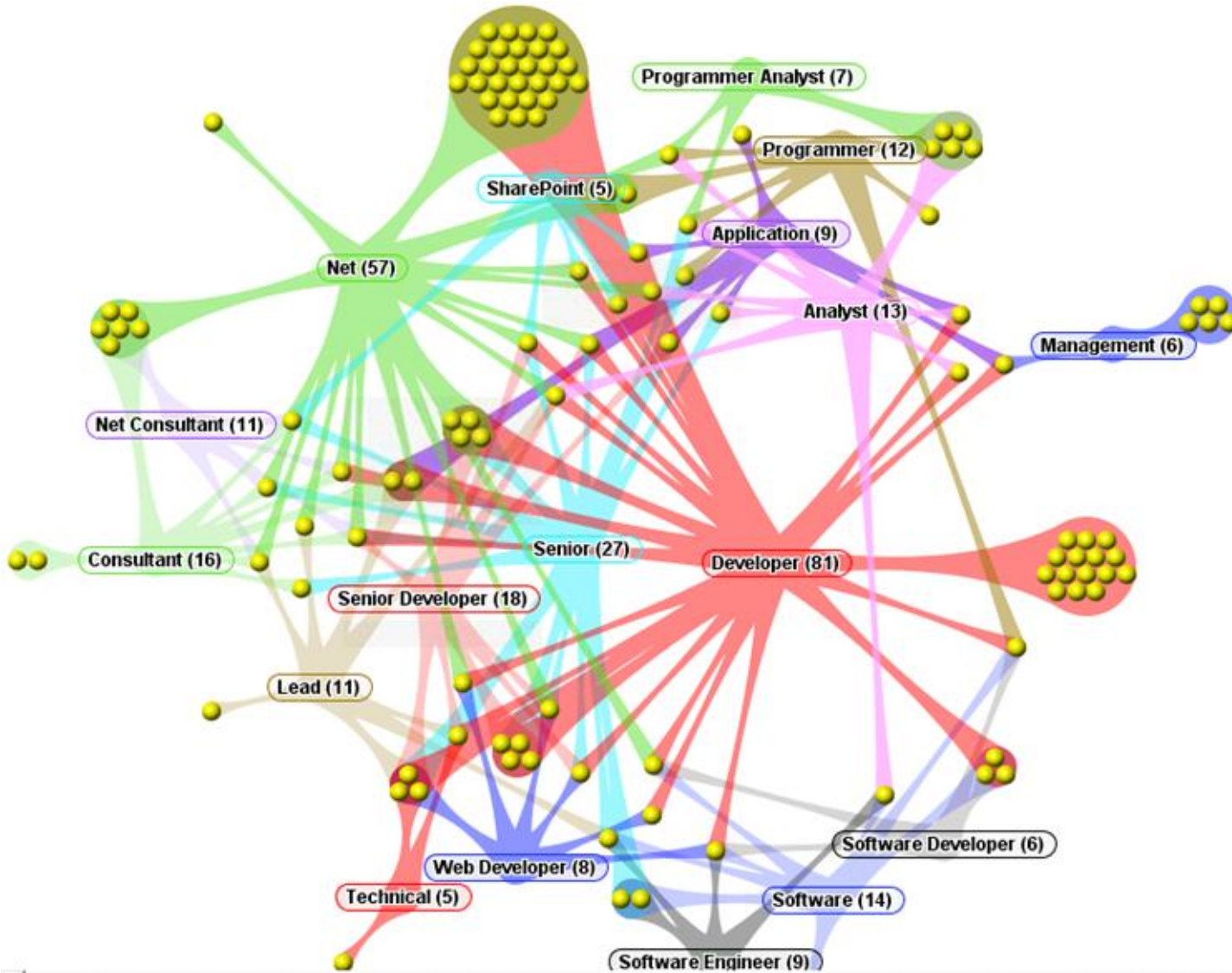
▼ Salary by education ** Figures may not add up to total due to missing or unavailable values*

Education	# Reporting	Avg.	Salary Distribution (percentile) ⓘ				
			10th	25th	50th ★	75th	90th
High School	1.00 %	\$64,572	\$54,908	\$56,160	\$58,240	\$68,640	\$75,712
Vocational	0.30 %	\$55,333	*	*	*	*	*
2-year Degree	78.10 %	\$64,831	\$50,000	\$55,000	\$62,400	\$72,800	\$81,120

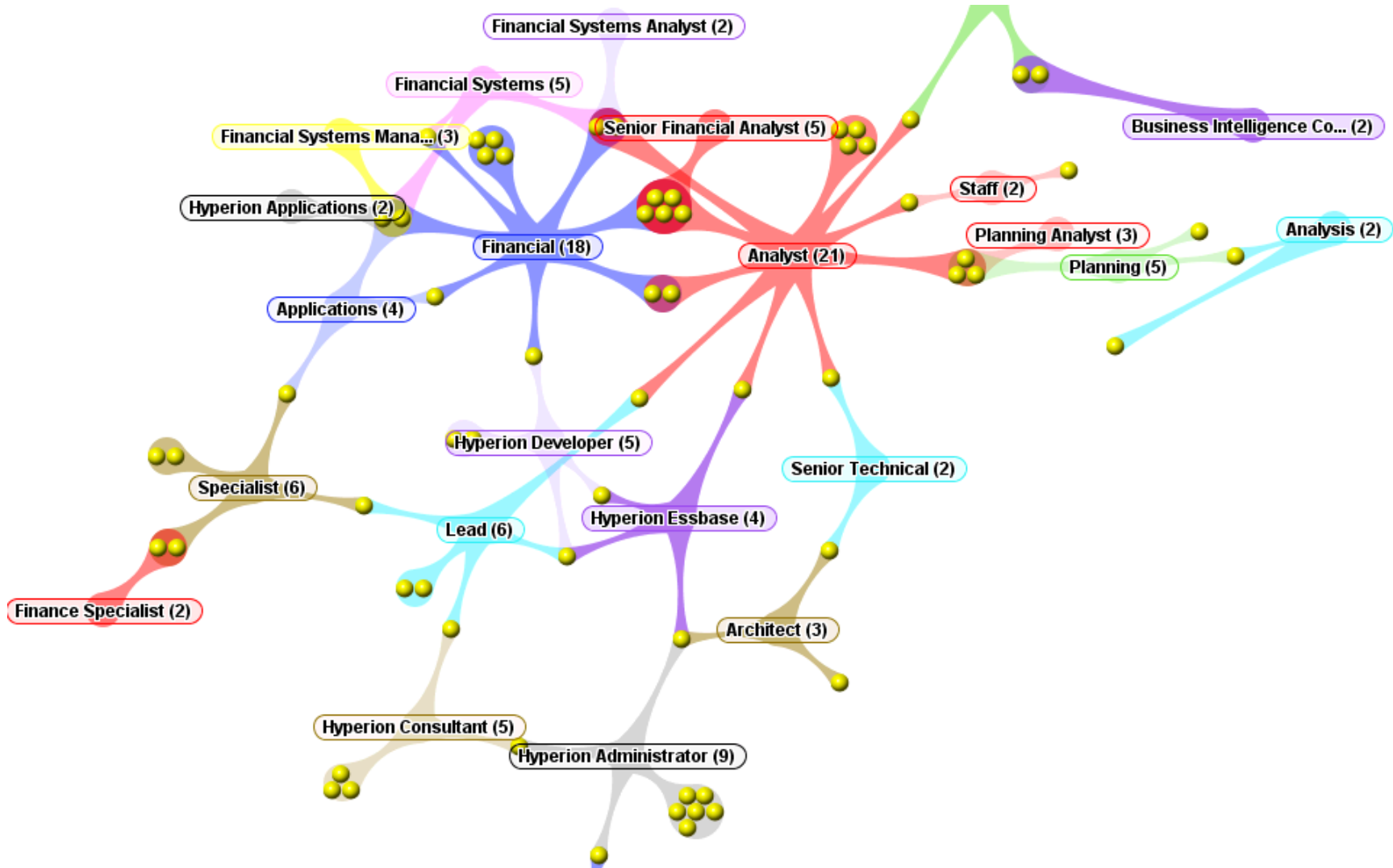
Clustering: Nursing



Clustering: .Net



Clustering: Hyperion Developer



Take Aways

- Know how your linguistics affect precision and recall and choose wisely; know how to tweak for your domain.
- A flexible software api that turn Solr into a SAAS type cloud app can greatly increase agility and adoption of search.
- Search isn't just about finding and navigating content... it can be used to learn from and create it, as well.

Contact

■ Trey Grainger



- trey.grainger@careerbuilder.com
- <http://www.careerbuilder.com>