A Look at Real-Time Labor Demand Through Online Job Postings

Real-time demand, a term used to describe measures of online job postings, has gained a great deal of attention as a method of identifying demand for workers that is more timely than "traditional" labor market information, such as Current Employment Statistics or Occupational Employment and Wages data. New Hampshire Employment Security's Economic and Labor Market Information Bureau has worked in a collaboration of eight states, known as the Northeast Consortium, to take a detailed look "under the hood" of real-time data.[1]

One of the Consortium's goals was to conduct research on new and existing sources of information — specifically, online job postings data — to provide up-to-date information that could be used for public and private decision-making. Given available products from different groups already using online job ads to identify demand, the Consortium had high expectations for usefulness of real-time demand data. After taking that look "under the hood," however, the Consortium became resigned to the fact that there are more data limitations than previously thought, and the validity of some data variables is highly questionable. In hindsight, this makes sense, since job postings are not written for data analysis. Research has helped define specific limitations of the data, providing data users a better understanding of what real-time data can and cannot do, and the potential for continued analysis of this data.

When evaluating real-time data there are two factors to keep in mind. First is the *universal* nature of online job postings. Regardless of how the data is extracted, there will always be certain limitations and biases inherent to online data. Secondly, there are *vendor-specific* issues to consider that relate to varying technologies and methodologies used by vendors in collecting and compiling online data.[2]

---

[1]. The Northeast Consortium partners include New Hampshire, Maine, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Georgetown University's Center on Education and Workforce, Burning Glass Technologies, and Direct Employers Association. The Consortium's research was funded by a State Labor Market Improvement Grant issued by the US Department of Labor's Employment and Training Administration.

[2]. Most vendors use computer technology called "spidering". The "spiders" find online job ads by scraping the internet in search of websites with job postings. Once a spider finds data, the data needs to be extracted and parsed into usable datasets, which is dividing text information into data categories. Vendors have different techniques for spidering, extracting, and parsing data.

Contents

Visit our Web site:
www.nh.gov/nhes/elmi

New Hampshire Online Job Ad Volume
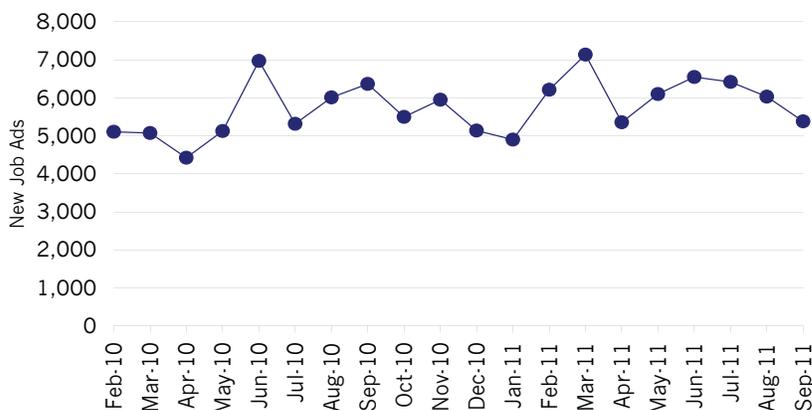
Note: Fluctuations in job ad volume over time may be attributed to change in demand and/or data collection processes.

Source: Burning Glass Data Parser 3.0.5 De-duplicated data.

One of the most critical conclusions was that an online job ad does not necessarily equal an open job, nor does it equal a hire. There are many reasons why a posting may not be an open job — there could be duplicate postings for the same job, the employer may not hire after placing an ad, there might be data captured incorrectly as a job ad, or postings could be for full time, part time, temporary, or one-time only work. Many reports use terminology like "total jobs" and "job counts" to describe the data, which perpetuates the notion that every ad equals a job. The term "job ad volume" provides a more accurate description.

When a job ad is posted, it frequently appears on more than one website, creating the potential for duplicate data. Vendors construct methodologies to identify and remove duplicated postings, known as a de-duplication process. In the Consortium's research, it was found that removing all duplicated postings is not possible, but the de-duplication process is able to reduce a significant amount of redundant data.[3]

Another discovery made through the Consortium's research was that data representing geographic areas smaller than an entire state was unreliable. This can be attributed both to the sheer lack of job location information in a large share of the online postings and to difficulties in separating geographic location from the job ad in the parsing process. Also, a minimum of one month of data was needed to reduce daily or weekly spikes, and longer time frames improved data validity. Analysis of online postings data is still fairly new and has no standardized process. To date, the Consortium has found that

Top NH Posting Employers 3rd Quarter 2011	Job Ad Volume
Liberty Mutual	319
Dartmouth College	284
Dell	259
Dartmouth-Hitchcock Medical Center	191
HCA Capital Division	160
Lowe's Companies, Inc *	152
Fidelity	144
C&S Wholesale Grocers, Inc	126
Citizens Financial Group	122
Omni Hotels & Resorts	114
TD Bank	111
Dell Perot Systems	102
Genesis Healthcare Corporation	99
Pizza Hut	88
Petco	72
BAE Systems	69
Technical Needs	68
Crossmark	66
Calm Water Business Partner	60
JP Morgan Chase Company	60
Brookstone Incorporated	59
Secureworks Incorporated	58
Walmart/Sam's	58
Bank of America	57
Staples	53
Catholic Medical Center	52
Sears, Roebuck & Company	50

* Lowe's job postings were verified. Job ads were prior to Lowe's announcing the closing of three New Hampshire-based stores.

detailed data review by an analyst is necessary to develop reliable information.

Here in New Hampshire, research was conducted using large data samples to determine the availability and validity of information spidered and parsed from online job ads. The data has some promise for identifying some of the skills, certifications, and educational requirements sought by employers; but warrant caveats as well.

Employers

Not all employers post jobs online. Some employers prefer to remain anonymous, and do not include their identity in a posting. When an employer name is listed, it can be phrased differently in every ad. For example, UPS can be listed as UPS, United Parcel Service, or United Parcel Service of America, Inc. In order to produce job ad volume reports by employer, the names must be standardized. Often, a human eye is needed to determine if the standardizations are valid.

Skills and Certifications

The full skill set and certifications required for a job are typically not included in job postings. Skills can have more than one meaning, be taken out of context, or be missed altogether by parsing software. The most valuable skills and certification data was related to technical skills, such as knowledge of specific software packages. Currently, such analysis is best referenced as a point in time, and trending over time should be avoided.

Educational Requirements

This information is lacking in a large share of online job postings. When requirements were specified, they tended to be higher than the average educational level of the broader workforce. Third Quarter had 17,829 postings, 56 percent with no educational requirement. Of the postings with requirements, 57 percent required higher than a Bachelor's degree.

Equally interesting, in a different study, among postings with no requirements listed, many were in professional positions for which

[3.] Data Quality Review, Burning Glass Parser 3.0.5

Education In NH Job Ads 3rd Quarter 2011	Share of Job Ad Volume
Not Listed	55.6%
Masters and above	2.4%
Bachelors	23.3%
Assocoate or High school	3.7%
High School/ GED	15.1%

a level of education could be assumed. Occupations like medical doctor, dentist, accountant, nurse, or computer-related positions are certainly likely to need education, and higher education to boot. Yet it was not included in job postings. Job ad writers may assume that applicants know what education is needed, so requirements are not listed.

Industry and Occupational Identification

Currently, it is extremely difficult to track data by standard industry or occupational codes. Job postings rarely contain industry specification and almost never include an industry code. Postings often lack necessary variables such as employer information that might be used to identify an industry.

Determining an occupation for a job ad was possible, particularly for occupations with a high volume of postings, but many job ads lack the essential information necessary for occupational coding leaving room for error. Job titles were available in most job postings and appeared to be quite accurate.[4] Grouping job titles by related job family (two-digit Standard Occupational Classification) has been the most effective method of analyzing occupations to date. Job titles may prove more valuable to identify emerging trends in online job ads rather than occupational classification, as job titles may be more relevant to job seekers and career counselors.

The Northeast Consortium had great hopes for the value of real-time demand data. Many of the expectations, however, were not fulfilled. Job ads posted on-line are prevalent in some occupational areas and non-existent in others. The postings are usually vague in geographic location details, and frequently do not include information such as salary,

educational requirements, experience requirements, and required skills. At present, the data has little value for steering public policy.

But this information is particularly indicative of rapidly changing trends in technology when used in point-in-time contexts. It could also be a useful tool for employment counselors, identifying employers with high posting levels for select job titles. Some occupational areas are well represented in the data, offering good insight into the market for those fields with much less of a lag time than traditional labor market information. When used in conjunction with other data sources, real-time demand data can be a useful addition to the larger picture of the labor market.

Erica Dickinson

Top NH Posted Occupations 3rd Quarter 2011	Job Ad Volume
Sales Representatives, Wholesale & Manufacturing	699
Retail Salespersons	652
Registered Nurses	583
Mechanical Engineers	566
Computer Software Engineers, Applications	314
Customer Service Representatives	302
General and Operations Managers	302
Computer Programmers	284
Truck Drivers, Heavy and Tractor-Trailer	283
First-Line Supervisors/Managers of Retail Sales Workers	281

Top NH Job Titles 3rd Quarter 2011	Job Ad Volume
Registered Nurse	369
Physical Therapist	226
Teller	214
Software Engineer	174
Sales Representative	162
Occupational Therapist	161
Sales Associate	140
Customer Service Representative	131
Drivers	123
Administrative Assistant	121

The Top Occupations table shows job postings that were auto-coded into Standardized Occupational Classification (SOC) coding; compared to the Top Job Titles table, which reflects the actual wording used in job postings

---

4. Based on data parsed by Burning Glass Technologies.