The Visualization and Analysis of the Online Postings of Information Technology Jobs

Syed Muzamil Basha [1], N.Ch. Sriman Narayana Iyengar [2*]

Abstract: Nowadays, gaining a substantial knowledge of the current job trends before obtaining a specialization in a certain field can be essentially important. Knowledge of the current job trends will help prepare the students before getting employed in different fields in the industry by training them with the correct and latest technologies that are required to do their jobs. This study aims to determine the most frequent and popular Information Technology job titles that can help prepare the future workforce. The data is analyzed from the job postings dataset through the KNIME miner, and then documents are created using Part-Of-Speech (POS) tagging through the Stanford Tagger. The documents are then preprocessed, and the most frequent and popular jobs are visualized using the Tag Cloud. The results showed the twelve most frequent and in-demand Information Technology job titles.

1. Introduction

Currently, students of Information Technology (IT) related courses acquire the very basic knowledge and skills that they may need when applying for jobs in IT-related industries. They learn general programming, engineering, and development methods, while the industry requires specializations in specific job positions. In several years of staying in universities or colleges, they will acquire the basics of at least five programming languages, basic concepts in networking, graphics design, and software development.

For students who graduated only with minimum qualifications to get jobs, knowing what the industry needs and acquiring the skills required to perform in their future jobs is essentially important. Understanding the current technologies that are needed to be accepted in job applications can be the best way to have an advantage in entering industries.

[1] School of Computer Science & Engineering, REVA University, Bangalore, Karnataka, India
  Email: muzamilbasha.s@reva.edu.in
[2*] Department Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telengana, India
  Email: srimannarayanach@sreenidhi.edu.in (Corresponding Author)

In addition, most graduating students have encountered problems in their interviews. Thus, some tips were discussed for cracking formal interviews and answering tricky questions.

Tip 1: Ask for clarification if the question seems unclear.

Tip 2. Make use of flexible terms for one's own advantage.

Tip 3: Attitude matters as much as your answer.

Tip 4: Practice control and pauses effectively.

Tip 5: Choose to answer a question with a question.

Tip 6: Always leave on a positive note.

This study will allow the listing of top twelve job titles from job postings made between 2004 and 2015. This list will be helpful for students and those who are looking for jobs to prepare themselves with the basic requirements and skills that the industry requires. The study makes use of KNIME data analytics platform in visualizing the top twelve job titles. This list will allow the students and industry stakeholders to prepare the future workforce related to IT.

The rest of this paper is organized as follows: Section 2 outlines the overview of the KNIME platform; Section 3 describes the design of the analysis using KNIME. Section 4 expresses the results achieved in the analysis. This paper is concluded in Section 5 with discussion on the contribution with limitations and future research direction.
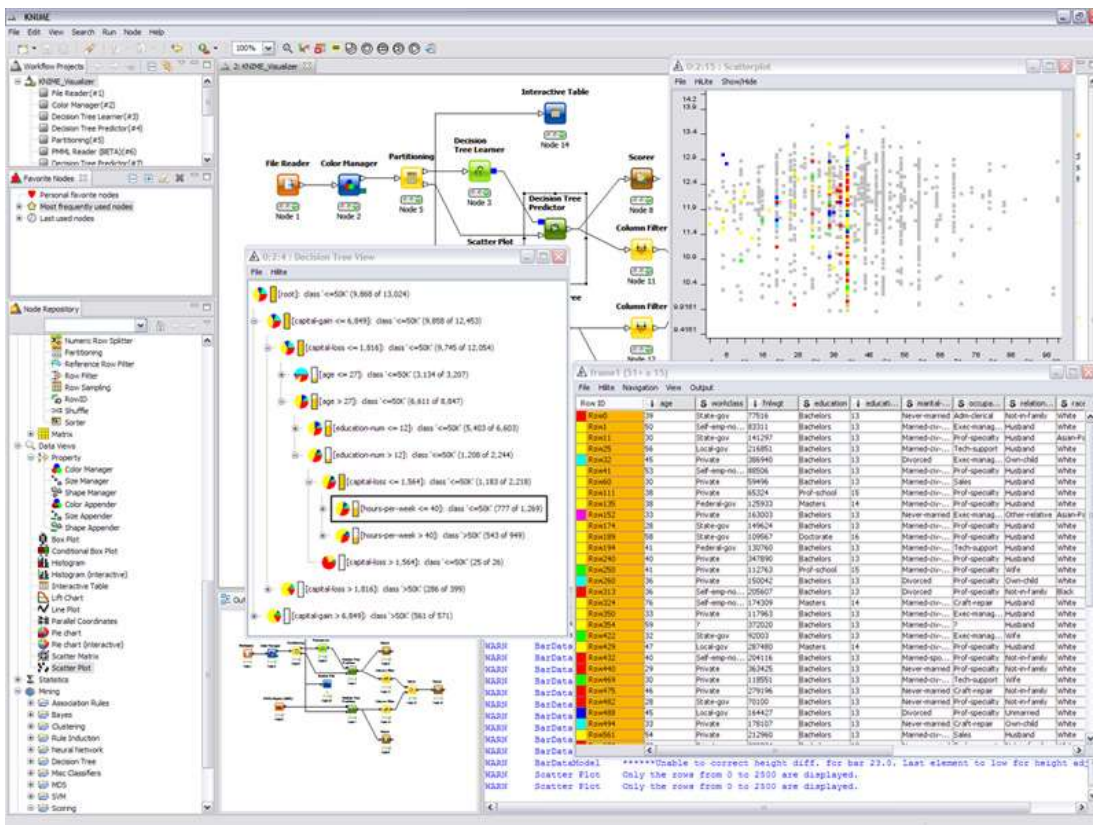
2. KNIME Overview

Figure 1. The KNIME Platform

KNIME, or the Konstanz Information Miner (KNIME) [1][2], which was started in 2004, is a free and open-source data analytics, reporting, and integration platform, as shown in Figure 1. The various components for machine learning and data mining were integrated through the modular data pipelining concept of KNIME. It allows the assembling, preprocessing, and blending of different datasets to provide modeling, analysis, and visualization. In addition, it allows users to visually create data flows, also known as pipelines, execute the analysis steps, and inspect the results, models, and interactive visual views. KNIME applications include pharmaceutical research, customer relationship management (CRM) customer data analysis, business intelligence, and financial data analysis [3]. An example workflow for KNIME is shown in Figure 2.
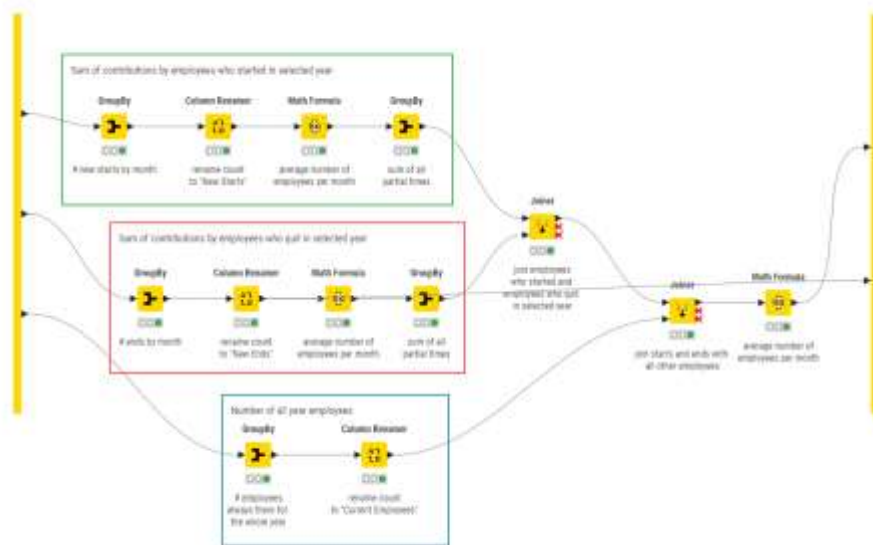

Figure 2. The KNIME Workflow [4]

The KNIME core version includes hundreds of modules for data integration (*e.g.*, file I/O, database nodes supporting all common database management systems through JDBC or native connectors: SQLite, SQL Server, MySQL, PostgreSQL, Vertica, and H2), data transformation (*e.g.*, filter, converter, splitter, combiner, and joiner), as well as the commonly used methods of statistics, data mining, analysis, and text analytics [5]. The capabilities of KNIME include the following:

- Drag-and-drop functionality to visually create complex workflows.

- The power of Pervasive DataRush performance and scalability is wrapped in KNIME nodes.

- GUI interface to all of the Pervasive DataRush Core Analytics Library.

- It is easy to download and install in any KNIME installation based on the 2.1.2 release.

- Integrates easily with all standard KNIME nodes and KNIME extensions.

- KNIME's core architecture allows the processing of large data volumes that are only limited by the available hard disk space.

- Additional plugins allow the integration of methods for text mining, image mining, and time series analysis.

- • KNIME integrates various other open-source projects, *e.g.*, machine learning algorithms from Weka, H20, Spark, the R project, and LIBSVM [6].
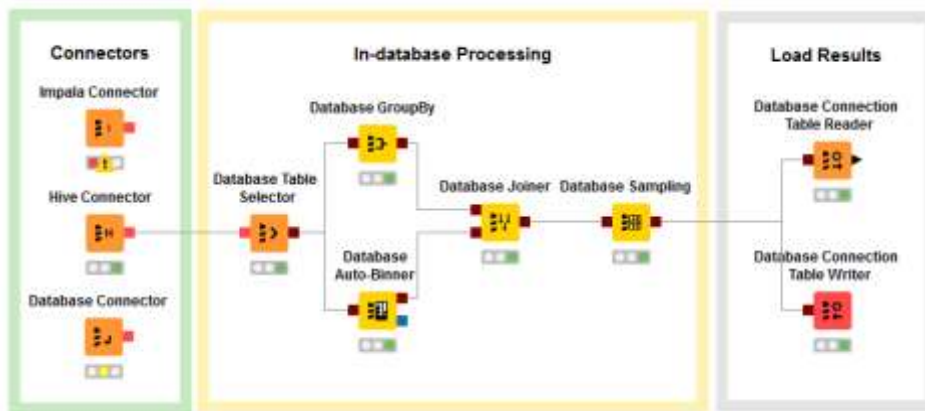

Figure 3. The KNIME Big Data Connectors

KNIME is implemented in Java but also allows for wrappers calling other code in addition to providing nodes that allow to run Java, Python, Perl, and other code fragments [2]. Figure 3 shows the KNIME big data connectors [7]. This presents a wide range of extensions used for integrating specific functionality like text processing, databases, hives, third-party applications, popular machine learning libraries, and more.

3. Methodology

Figure 4. Job Postings Dataset from kaggle.com

The KNIME table reader node was utilized to easily perform the data import. In this regard, the data set of job postings can be organized carefully. The dataset used in this study was collected from Kaggle datasets [8], as shown in Figure 4. It consists of job postings on the Armenian human resources portal Career Center between 2004 and 2015. In Figure 4, the rows consist of different job postings, while the columns provide the descriptions of each of the job posts. This job postings dataset will be the input in the experiments for this study in terms of visualizing the top jobs as well as understanding the skills and minimum requirements needed by students who wish to apply for these jobs.
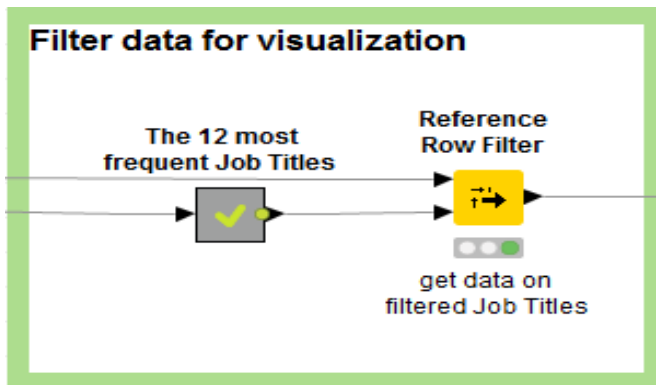

Figure 5. Filter Data for Visualization

The output from the KNIME table reader node is filtered and visualized, as shown in Figure 5. The top twelve job titles are grouped based on the most occurrence of job titles, and the same is sorted in descending order. Each string in the job postings dataset is considered a document, and a part of speech tags are attached to it using the Stanford tagger that is available in KNIME. Then, the documents are preprocessed to make the dataset readable and ready for analysis. In the preprocessing stage, the punctuation marks are removed, and the numbers and stop words are filtered out. After filtering out and data cleaning of the preprocessing, the lemmatization process is carried out using the Stanford Lemmatizer. The output from the Lemmatizer is now considered the preprocessed data.

The next stage will be the preparation of tags for visualization. In this stage, only the proper nouns (NNP, NNPS) are considered as input in creating the bag of words (BoW) for the corpus analysis. Then, the relative term frequency is calculated to find out the most frequent words from the job postings, as shown in Figure 6.
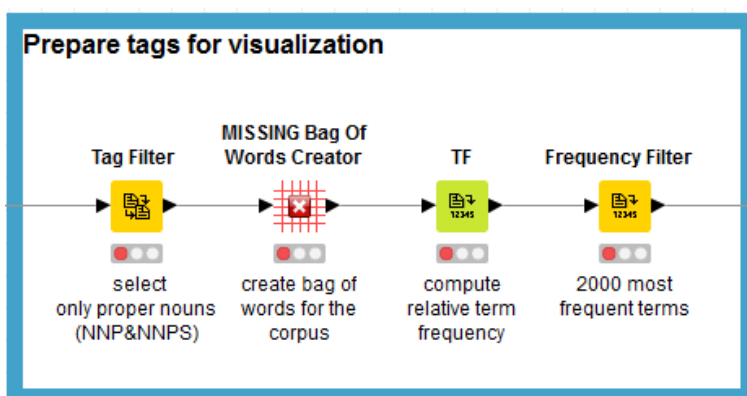

Figure 6. Preparing Tags for Visualization

4.   Results and Discussion

The results of the visualization from KNIME have shown the twelve most influential job titles related to IT, which are listed as follows: graphic designers (GD), Java developers (JD), programmers (P), senior software engineers (SSE), software developers (SD), software engineers (SE), web developers (WD), IT specialists (IT-s), PHP developers (PHP-D), senior QA engineers (SQA), senior software developers (SSD), and senior Java developers (SJD). Figure 7 shows the top 12 most influential job titles based on the analysis of the job postings dataset.

Figure 7. Top Twelve Most Influential Job Titles Related to IT

The results show that, in 2004, the most influential job title was programmer. They are most needed in designing the core parts of interfaces between the user and the computer (*i.e.*, human-computer interfaces (HCIs) and user interfaces (UIs)). Whereas, in 2005, the most influential job title was software developer. This year, they are playing an important role in designing and developing service-related systems and software. It was in 2005 that the use of information systems that provide automated services to users grew rapidly.

The results also show that in 2006, a new field called web development came into play as the number of Internet users significantly rose. This year marks the emergence of Web 2.0, where the web has advanced significantly from static to dynamic and from passive to active. In this regard, web developers were the most influential job title in 2006. In 2007 to 2008, Java developers had more demands and job postings as the Android OS, an open source operating system, was released. Java developers are needed to program Android devices as well as to develop Android apps.

In 2009 to 2011, the recession in the field of Information Technology drastically slowed the number of intakes. Again, from 2012 to 2015, a significant number of web developers were needed for the IT industry.

5. Conclusion and Future Works

This paper has analyzed the job postings related to Information Technology (IT) to list the most influential job titles that will be helpful for college students in preparing for their journey before their employment in different industries. The results showed that jobs in the IT industry are completely independent of past trends and completely dependent on current industry needs. In this regard, students must focus more on developing their skills in web development and Java programming, as these are the minor requirements based on the most in-demand job titles listed. Equipped with these skills, they will have an ample number of opportunities to get the jobs in the IT industry.

In the future, a study on identifying the skills needed to inculcate among student communities will be addressed.

References

[1]   KNIME, "*KNIME Analytics Platform*", www.knime.com/knime-analytics-platform (November 12, 2018).

[2]   KNIME, "*Getting Started Guide for KNIME Analytics Platform*", www.knime.com/getting-started-guide (November 12, 2018).

[3]   KNIME Hub, "*Welcome to KNIME Community Hub*", www.hub.knime.com/ (November 12, 2018).

[4]   Wikipedia, "*KNIME*", www.en.wikipedia.org/wiki/KNIME (November 12, 2018).

[5]   KNIME, "*KNIME Components Guide*", www.docs.knime.com/latest/analytics_platform_components_guide/index.html#introduction (November 12, 2018).

[6]   Pervasive DataRush, "*Pervasive DataRush for KNIME*", www.pervasivedatarush.com/products/Pages/DataRushforKnime.aspx (November 12, 2018).

[7]   KNIME, "*KNIME's Open Ecosystem*", www.knime.com/knime-big-data-connectors (November 12, 2018).

[8]   Kaggle, "*Online Job Postings*", www.kaggle.com/madhab/jobposts (November 12, 2018).