

TASK REPORT

Approach:

Resumes were extracted, parsed and segregated into the following main categories-

- Category
- Education
- Skills
- Experience
- Achievements
- Others (Awards, Publications, Certifications, Volunteer Experience, Interests, Languages, Licenses, etc.)
- Summary

Segregation was performed by identifying the different headings in each resume using a list of synonyms for each category. This helped to get a standardized extracted_resume dataset where each resume was identified by its File ID (stored as ResumeID), in order to be used in the next stage of Resume-JD matching.

The given Job Description dataset was filtered at first to remove the ones without having any Educational or Skill requirements mentioned in the JD. 15 job descriptions were then randomly sampled from the cleaned dataset.

For tokenizing and creating embeddings of extracted texts of education and skills in resumes and job descriptions, latest **Sparse Lexical and Expansion (SPLADE)** model was used. This gives more robust and learnable sparse embedding by enhancing it with the capabilities of a pretrained language model like BERT can identify connections between words/sub-words.

The CV-JD matching was divided into 2 parts that were weighted equally to get the final cosine similarity scores-

- Educational Requirements of JD ↔ Educational Qualifications of Applicant in Resume
- Required Skills of JD ↔ Skills of Applicant in Resume

For each of the 15 JDs selected the entire Resume dataset was traversed

Challenges faced:

1. Different resumes have different heading names and have random ordering
2. Work experience of most resumes had multiple positions written across many pages, this made it very hard to detect the entire experience section. This also meant it couldn't be leveraged properly without losing out on imp information

3. There were many other additional headings that could provide better insight of the applicant's experience, knowledge and skills, but weren't present in most resumes
4. There were many non-unicode characters and improper whitespacing
5. Headings weren't always prominently written
6. Many JD didn't have any educational requirements or required skills specified
7. Educational qualifications have varied names and formats of specification, they also have different levels of importance that can't be standardized
8. Lack of labeled data meant I couldn't use DL
9. Vocabulary mismatch problem: There's a typical lack of overlap between Job descriptions and resumes, because often contain ambiguous language and jargon specific to certain industries or roles. Since humans have different ways to write and express the same thing, we can't perform simple phrase or named-entity matching to find best matching resume for a particular JD.

Solutions:

1. Used list of synonyms for different heading names and made the code order-agnostic
2. Work experience was tackled differently so that all the text that followed the 'Experience' heading was collected together until the next heading, in order to capture all the multiple positions held by the applicant
3. Sections like Summary, Achievements, etc. couldn't be leveraged due to lack of presence in many resumes, which could potentially bias the matching algorithm against them
4. Entire text extracted from the PDF had to be Unicode normalized to remove any other special characters, also extra whitespaces (' ', '\n', '\t') were removed.
5. Resume headings must be written prominently with bigger/bolder font with newlines before and after it. So to tackle this we modified the heading-matcher algorithm so that it works even if headings are not formatted differently from the rest of the text.
6. The job descriptions were filtered to remove the ones not having any Educational or Skill requirements, only then 15 JDs were randomly sampled from the filtered dataset.
7. Using fuzzy or regex matching was not enough for this purpose. On the other hand, popular language models like BERT or T5 focus more on the semantic meanings of texts, and loses the importance of the actual degree levels. Hence SPLADE embeddings were used to find whether the applicant's qualifications satisfy the requirements.
8. Training spacy or BERT or T5 models was not possible, hence a much more efficient embedding model SPLADE was used
9. SPLADE model outperforms BERT, BM25, TF-IDF models because it can leverage both keyword-matching to give importance to phrases in skill or education section like the level of education acquired and also capture semantic meaning of a text.

Top 5 candidates for each job description based on similarity scores:

Let's consider one of the JD among the 15 used for matching-

company description

project find's mission is to provide low and moderate income and homeless seniors with the services and support they need to enrich their lives and live independently

today project find operates three supportive housing residences that are home to about 100 people and four senior centers with over 1,000 members

our members and residents range from the healthy and active to the frail and homebound to the homeless

for all these individuals project find is a critical resource providing housing meals and programs that help individuals navigate the challenges of aging by encouraging community engagement and healthy living

please visit our website at

projectfind.org

job overview

as the director of finance you will be responsible for building and enhancing the financial infrastructure of project find

the director of finance is also responsible for all aspects of the financial operations for project find

reporting directly to the executive director the director of finance provides accounting budgetary operational and programmatic support to all agency programs and activities

the director of finance directly oversees and supervises the financial accounting team

they will serve as the primary agency contact for all funding contractual matters proposals for new or extended programs and new business development

the director of finance is seen as a thought leader and partner who knows how to manage the inevitable rigidity between business objectives and resource utilization

key responsibilities

provide insightful well informed and trusted organizational counsel to the executive

director executive team and the board on both financial and broad based business matters

develop internal tools and systems to provide critical financial information to help

accelerate growth and profitability and make actionable recommendations on financial goals

and strategy track and ensure accurate and timely reporting of project find's financials to

board government contracts sources and donors oversees submission of vouchers for payment

from a variety of contract sources oversees annual budgeting process for agency and housing

corporations assess organizational performance against both budget and company's long term

strategy manage the corporate financial planning budgeting and forecasting function of the

agency proactively engage leadership team to develop short medium and long term financial

plans and projections lead and develop the finance team develops and refines policies and

procedures manual for finance department works with board and financial advisors to steward

investment fund other duties as needed based on organizational needs qualifications

bachelor's degree from an accredited college or university required masters degree preferred
cpa is a plus

years of senior management experience with organizations with operations in excess of

million per year and over 100 employees in a nonprofit agency recent and relevant experience

as a finance director at a nonprofit agency with state and federal funding portfolio with

strong familiarity with software revenue recognition metrics and best practices

proven track record as a results driven supervisor

prior experience serving as a key member of a senior executive team with the ability to

provide and receive input from ceo board and management peers regarding the financial

implications of all major business decisions and strategic objectives excellent business

acumen and financial operational orientation with strong fpa skillset experience in a

regulated industry preferred nonprofit experience is highly preferred and a strong plus

outstanding analytical and problem solving skills able to communicate effectively at all

levels of the organization including directly with the board maintained working knowledge

and understanding of regulations laws and industry best practices to ensure compliance with

all applicable federal state and local financial laws ability to interpret appropriate laws

and policies and ability to advise management and employees accordingly

excellent verbal written communications and detailed oriented skills

highly organized detail oriented and able to handle multiple projects simultaneously in a

fast paced environment must be able to travel between sites as needed

associated topics

associate director chief investment officer director finance financial director general

operations manager manage operation president finance recruit vice president

Here are the top 5 matches found for the above JD-

index	ResumelD	Category	Education	Skills
9	12802330	ACCOUNTANT	BS : Accounting Business Administration , 2005 University of Maryland University College i 1/4 City , State Accounting Business Administration CPA Candidate	accounting, Accounting Systems, ad, analytical skills, auditing, automation, billing, budgeting, budgets, Budget, business operations, business process, Cash-Flow, cash flow, closing, contracts, CPA, Crystal Reports, Database Management, Essbase, senior management, financial, Financial Accounting, financial analysis, financial analysis/planning, Financial Reporting, financial statements, fixed assets, Forecasting, general ledger, Great Plains, Hyperion, inventory, ledger, law, market, MAS 90, Excel, MS Office, negotiation, Enterprise, Peachtree, personnel, policies, pricing, processes, quantitative analysis, QuickBooks, reconciling, relationship management, reporting, researching, Research, sales, Sarbanes- Oxley, Strategic Planning, Tax Preparation, tax, taxes, Trend
1252	95792386	CONSULTANT	Master of Science : Accounting/Taxation San Francisco State University Accounting/Taxation Extensive course training in the fields of corporate finance, taxation, and business law. Bachelor of Science : Business Administration Accounting San Francisco State University Business Administration Accounting Coursework in accounting, finance, marketing, and information systems.	Accounting, accountant, Accounts Payable, Accounts Receivable, streamline, application design, auditing, billing, book-keeping, budgeting, budget, business law, business process, Cisco, closing, excellent communication, interpersonal skills, contracts, corporate finance, Cost Accounting, cost analysis, CPA, client, clients, data analysis, database, delivery, documentation, DOORS, due diligence, senior management external audits, finance, Financial, financial planning, financial reporting, financial reporting, financial statements, forecasting, foreign exchange, functional, General Ledger, general ledger accounts, information systems, Inventory, managing, marketing, meetings, enterprise, Networks, Oracle, Oracle Application, Oracle Financials, Oracle Financial, Platinum, policies, problem solving, publications, quality assurance, report writing, reporting, requirement, revenue recognition, sales, Sarbanes-Oxley, Supervisor, system analyst, tax, tax research, tax returns, team player, trend, upgrade, validation, variance analysis, year-end Activities and Honors California CPA. Completed the UC Santa Cruz Certificate of Project and Program Management. Coursework in SQL and Relational Database. 3
1683	84356308	FINANCE	Bachelor of Science : Mathematics 1997 University of the Punjab - City Certified Public Accountant -CPA : Accounts, Audit and Finance 2012 Institute of Certified Public Accountants - City	Accounting, Accounts Payable, Accounts Receivable, Acquisitions, Administrative, Ad, Auditing, Budgeting, budget, client management, Contracts, Contract Negotiation, Corporate finance, CPA, client, clients, clients i, decision making, Due Diligence, ERP, senior management, features, Finance, financing, Financial, financial operations, Financial Planning, financial reports, Financial Reporting, financial statements, Forecasting, GL, HR, insurance, international business, Investment Management, leadership, managing, management reporting, Market, meetings, Mergers, MS Excel, negotiations, next, PAGE 2, Performance Analysis, pricing, problem solving, profit, Project management, quality, reporting, retail, Risk Management, sales, Strategic, strategic planning, tax compliance, tax, Treasury, Valuation
1687	88038965	FINANCE	Bachelor of Science : Accounting University of South Carolina i 1/4 City , State Accounting Certified Public Accountant (CPA) State of South Carolina 3	Accounting, audit reports, billing, budgets, budget, closing, cost accounting, Certified Public Accountant, CPA, Decision Support, direction, senior management, Finance, financial, financial analysis, financial audits, financial operations, general ledger, Hospice, internal audits, Lawson, leadership, managing, Office, modeling, monthly financial close, Payroll, processes, progress, Sarbanes-Oxley, staffing, strategy, supervisor
1690	93653247	FINANCE	Master of Business Administration , Business Economics/Finance State University i 1/4 City , State Business Economics/Finance Bachelor of Science , Finance/Accounting Hoboken State University i 1/4 City , State Finance/Accounting	Budgets, budget, business analysis, contracts, Contract Management, Financial, forecasting and financial, financial operations, functional, leadership, mentor, Office, enterprise, optimization, organizational, performance reviews, processes, Program Management, Project Management, project planning, project plans, quality assurance, quick, reporting, strategy, strategic analysis, unique, website development

Recommendations/ Insights:

- Nearly 3/4th of the Job description dataset provided had no educational or skill requirements specified, so they couldn't be utilized for matching
- It would be better if we could get some annotated datasets for training spacy or NER models to improve the resume extraction
- Also if we had a large labeled dataset for resume-job matching scores then we could have trained CNN/BERT/T5 models for further enhancement of the matching algorithm.