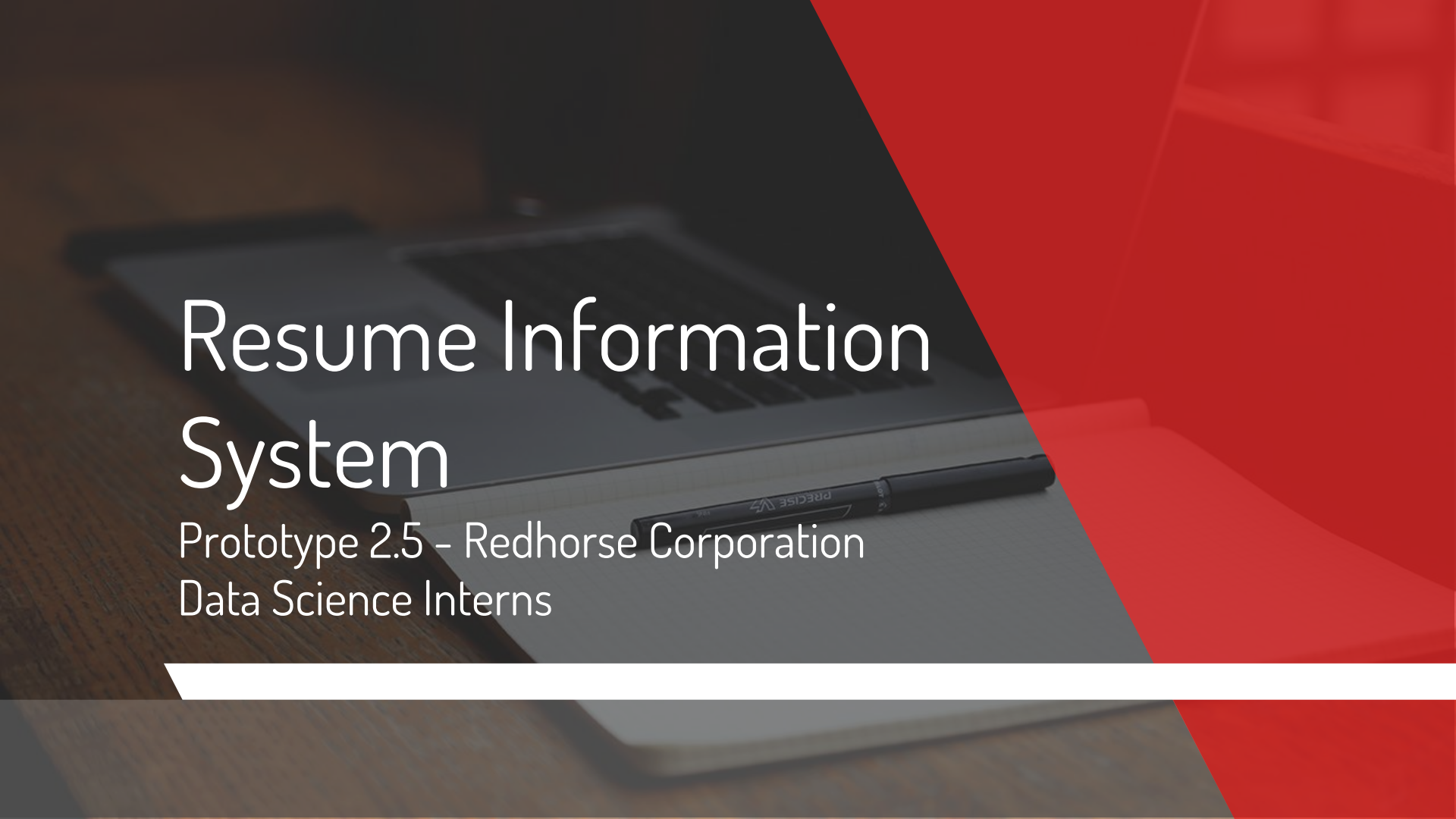


Resume Information System

Prototype 2.5 - Redhorse Corporation

Data Science Interns

Difficulties to be Solved

- **Identifying qualified applicants** and/or employees for a specific project is an essential, yet currently **arduous**, process
- Collecting, managing, and filtering resumes by hand leaves little time to devote towards understanding individuals on a **more human level**
- Redhorse Corp. requires centralized **in-house tools** to manage data and its integrity related to:
  - Current and Past Employees
  - Applicants
  - External Job-boards

System General Overview

- Utilized by HR and other Redhorse employees to expedite **targeted search** process of potential job and contract candidates
- To search, user enters certain **required skills** and **preferred skills**
  - **Evaluates** and returns only candidates that meet required skills threshold by evaluating resume text
  - **Ranks** candidates based on amount of required and preferred skills present in each candidate's resume
- User can additionally filter candidates based on associated **categories**

System Functionality

Build and deploy an in-house resume database system with the following functionality:

- **Targeted Searchability**
  - Relevant Skills
  - Domain Area
  - Clearance Level
- Data Security
- Periodic Updates
- Candidate Traceability
  - Redhorse Employee
  - Previous Employee
  - External
- **Candidate Summarization**
  - Keywords
  - Concepts
  - Categories

Feedback of Generation 1 System

- Update database nightly
- **Wildcard search ability**
- Mine job boards for potential candidates
- Determine if skills are ‘in progress’
- **Display if a person is an employee or not an employee**
- Consider Entity Resolution Tool
- **Test over large number of resumes**

Improvements Incorporated into Generation 2.5 System

- **Integrated Watson Natural Language Understanding**
  - extract related categories, concepts, entities, and keywords
- Integrated alternative **in-house Machine Learning algorithms**
  - Watson/Bluemix costly and black-box
  - Desired cheap/free alternative with greater fine-tuned control
- Display and search for candidates that have **associated security clearances**
- Added ability to **download resumes in bulk**
- Increased name recognition and PII extraction

Improvements Incorporated into Generation 2.5 System

- **Office 365 employee information integration**
- Increased ability to determine a current Redhorse Employee
  - Search option to display only employees or only applicants
- Input for **minimum number of required skills** to filter users
- Display percentage of required and prefered skills met
- Ability to **weigh or boost specific skills** (actively in development)
- OAuth 2.0 **secure login** using Redhorse outlook credentials

Technical Specifications

- **Elasticsearch** (Local)
  - Database
  - RESTFul Search Engine
- Front End UI
- Synonym Generation
  - Facebook fastText
- **Watson Integration** (Bluemix)
  - Natural Language Understanding
- **OAuth 2.0 Secure Login**
- Debian9/AWS Server
- Django Web framework
- Python 3.5+

Project Difficulties

- Knowledge of and access to pre-existing internal Redhorse systems
  - Resume **receivement, review, and storage** process
  - Egnyte, MyStaffingPro, and Redhorse Website
- Employee Metadata
  - **No centralized database** of relevant employee information
  - Updated **by hand** on Excel spreadsheets
- Access to external services and systems
  - **Built on Bluemix/Watson trial account**
  - Recently procured Amazon Web Service instance
  - Senzing 2.0 communication and underdeveloped API
- **Searching for open-source Watson and Machine Learning alternatives**
  - Bluemix/Watson expensive PaaS with questionable future
  - IBM cannot provide additional control and understanding of Watson (i.e. **Blackbox**)
- Open-source packages are powerful but not always well-maintained
  - Gensim and Rapid Automatic Keyword Extraction (RAKE) issued bug reports
- Limited access to Redhorse resumes
  - **Egnyte contains only ~185 resumes**

Future Features

System Feature	Description	Difficulties Foreseen
Automatic Data Updates and Resume Extraction	• Weekly/monthly updates of employee metadata from Office 365, Paychex, MyStaffingPro, etc. • Update Watson as it trains	• Employees to keep up-to-date records on Office 365 • Existence of accessible API and permissions • Watson requires a non-trial Bluemix account
Entity Resolution: Senzing G2 Tool	• Links entities that are the same and derives relationships	• Senzing is in beta stage and interfaces underdeveloped • Not compatible with Debian Linux distributions, must be run on Redhat/CentOS (less supported)
Open-source Watson Machine Learning Alternatives	• Provide entity, keyword, and category extraction without relying on black-boxed Watson	• Longer development times, but more flexibility, control, and portability • Requires additional technical research
External Job board API Integration	• Search and extract qualified candidates and/or applicants	• Applicant data exists on MyStaffingPro (can't extract) • No public API for ClearanceJobs
RFI Skills Extraction	• Extract related skills and concepts from categories and use RIS to find relevantly-skilled employees	• Need to perfect similar skills search • Need to perfect keyword & skills extraction

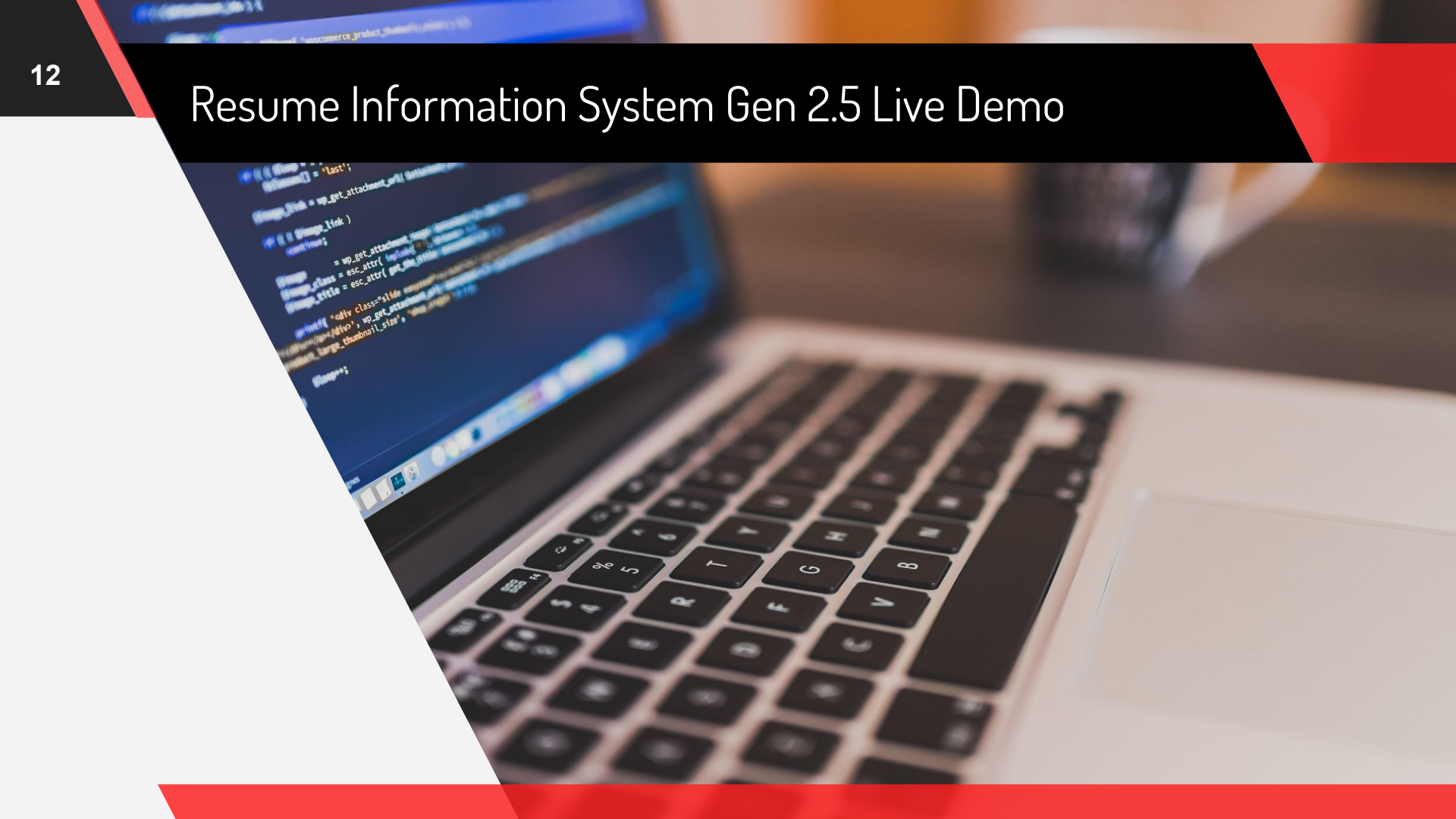
Transition to Past Performance Information System (PPIS)

"Although the current implementation is specific to resumes, RIS has the potential to be used in a variety of other systems that rely on **Natural Language Processing** of unstructured text documents. For example, one could store **Project Descriptions (PDs)** in a database rather than resumes. One could then quickly search over these PDs for specific skills, tasks, events, etc. Another example could include **Requests for Information (RFIs)** or related documents. One could store these documents and search for skills required for past or future Redhorse projects."

- Redhorse 2017 Data Science Interns

- **Info Required for Potential PPIS**
  - Project Descriptions
  - Individuals on project and roles
  - Skill of individuals
  - Domains/scopes for which Redhorse projects exist
  - Successful and unsuccessful bid's RFIs
  - Project success or failure
  - Digital formats (pdf, docx, etc.)
  - Project documentation/reports/forms
  - Centralized employee information database

Resume Information System Gen 2.5 Live Demo



Recap of the Resume Information System 2.5

- **Targeted Searchability**
  - Required & Prefered Skills
    - Minimum Skills Matched Filter
    - Candidate Scoring & Ranking
    - Thesaurus Lookup/Generation
  - Clearance Level
    - Secret, TS, TS/SCI
  - Redhorse Emp. Identification
    - Office 365 Integration
  - Domain Areas of Candidate
    - IBM Watson Generated
- **Watson Generated Summarization**
  - Keywords
  - Categories
  - Concepts
  - Contact Information
- **In Progress System Additions**
  - Customized Candidate Scoring
    - Weigh Individual Skills
  - Entity Resolution (ER)
    - Senzing 2.0 Integration
  - In-House Machine Learning Algorithms
    - Relevant Keyword Extraction
    - Unstructured Category & Concept Generation
  - Tailored Search Summary
    - Sharable Files for Candidate Info. Exchange
  - Office 365 Authorized Sign-In
- **Potential Future Additions**
  - Job Board Integration
  - Automated PDs to Skill Conversion

REDHORSE

Questions

Feedback is appreciated