Computers Assist Humans in Human Resources

Lance Tokuda

Resumes are the primary vehicle employers use to identify qualified job candidates. Thousands of companies across corporate America have used the recruiter to match tens to thousands of resumes to tens to hundreds of job opportunities. This chapter discusses Resumix, the world's first intelligent resume-processing system. Resumix reads applicant resumes, creates a precise resume summary, matches candidates to job openings, generates reports, and prints applicant acknowledgment letters with a bit-map signature from the appropriate hiring manager. Systems are currently installed and running at Sun Microsystems (1/89), Advanced Micro Devices (5/89), Bank of America (7/89), National Semiconductor (11/89), Sequent Computer Systems—Oregon (12/89), BPI (1/90), Digital Equipment Corporation (1/90), Amdahl (2/90), Texas Instruments (2/90), Advanced Micro Devices—Texas (4/90), General Motors (5/90), and AT&T (6/90).

Problems in the Corporate Human Resources Office

Resumix's Fortune 500 customers share two common needs: (1) an efficient method to process the hundreds or thousands of resumes received each week and (2) an efficient method for recruiters to search

for qualified candidates among the thousands of applicant resumes received.

Resume-processing procedures vary with each company. A typical human resource (HR) organization analyzes, sorts, copies, and distributes resumes to the appropriate hiring managers. Individual recruiters then apply their own filing techniques to store resumes for future use. Response letters are generated for each candidate and resumes are filed by job category, month received, and last name.

Candidates applying to one Fortune 500 customer who inquired about the status of their resume often received the response, "Please allow three weeks for our staff to process your resume, then call again if you do not receive a response." On occasion, a candidate's resume was lost in the paper shuffle. By the time this loss was confirmed, the job opening had already been filled, making the applicant angry. A surprising number of candidates, aware of this possibility, regularly sent multiple copies of their resume to this company. Finally, many excellent candidates accepted other positions before recruiters or hiring managers at the target company were even aware of the candidate's existence.

Once resumes are processed and stored, a new set of problems arises. Recruiters receiving a new job requisition typically have three options: (1) search through existing resume files for a match, (2) place an advertisement announcing a new opening, or (3) call on the services of a head-hunting agency.

In an attempt to save funds, the recruiter pulls out a stack of 400 applicant responses to previous ads for similar (but not the same) positions. For example, a Sun Microsystems recruiter searching for software engineers with C and Unix experience had to transport 2400 resumes received in the past 3 months if he wanted to take his work home. The advantages of an online computer search capability are apparent, but many recruiters do not have the time, patience, or astute HR assistants to code all incoming resumes by skill and manually enter the data in a database. In the past, vendors have offered optical character recognition (OCR) systems that provide a key-word text-search capability on all text documents. Unfortunately, key-word searches cannot distinguish the senior engineer with a Master of Science degree and VLSI Design experience from a first-year data-entry clerk who did some spreadsheet design on an MS-DOS computer for VLSI, Inc. Furthermore, these systems fail miserably when searching for candidates with C experience. Recruiters then resort to placing ads in journals or local newspapers costing thousands to tens of thousands of dollars. Others retain the services of executive search firms. Fees charged by these firms range from 20 percent to 35 percent of a hire's first-year salary.

Automating the Resume Flow

Resumix hardware consists of an optical scanner–character recognition unit, a workstation server, and zero or more workstation clients (figure 1). Resumix software provides a mouse-driven graphic user interface. Actions are executed by pressing graphic buttons. On a typical day, a clerk feeds resumes into the scanner at the rate of 250 each hour (assuming two-page resumes). When the scanning is finished, the clerk pushes the "process" button. Resume images are broken into text blocks and sent to the optical character reader for text recognition. The text blocks are then reassembled and stored with the image for retrieval by a recruiter. An extractor module applies a series of patterns and rules to the text. The following information is automatically extracted: last, first, and middle names; as many as two addresses (permanent, temporary); as many as three telephone numbers (home, message, work); degrees (level, major, year, school, grade point average); jobs (years, company, job title); and skills (for example, circuit design, relational database [or RDB or RDBMS], and C). Based on the extracted information, Resumix then categorizes an applicant in one or more applicable categories and stores this information with the extracted data in an applicant summary. When processing is completed, clerks compare the extracted information with a computer image of the resume to verify it. Candidates whose resume could not be properly extracted (usually because of a poor-quality resume copy) have their data manually entered or are asked to send a clean copy of their resume. After resumes are verified, the clerk pushes the "match resumes to open requisitions" and the "print response letters" buttons. The total elapsed time between the submission of a resume and the receipt of a response letter ranges from two to four days.

Recruiters receiving a new job requisition enter the requisition criteria into Resumix. Recruiters can search by category; extracted skills; degree; major; school; years of experience; company; date received (last week, last month, and so on); and location, as determined by the area code for the telephone numbers. The ability to place this information in a computer-searchable form is the key to the success of Resumix. Searches through stacks of resumes for a candidate with a unique set of skills are completed in seconds instead of hours or even days. All resumes that are received are considered, not just those filed by the individual recruiter. Many recruiters would take their stacks of resumes home for a midnight reading. "We're talking about a change of lifestyles (for these recruiters)," says corporate HR manager Stephanie Buchholz of Sun Microsystems. As new resumes are scanned into the system, the "match resumes to open requisitions" process creates a list

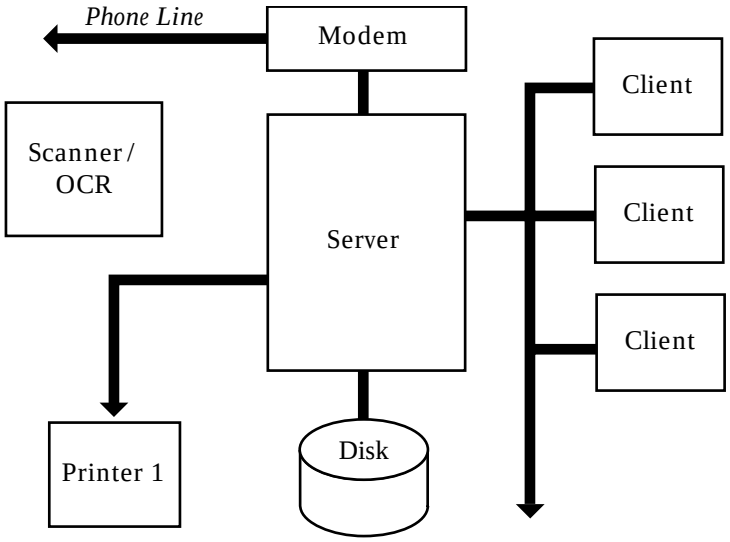

Figure 1. Resumix Hardware Configuration.

of matching resumes entered into the resume database since the last time the button was pushed. As a part of their daily ritual, recruiters push the "display requisition matches" button to view the resumes of new candidates matching the requisition. About the only thing Resumix doesn't do is telephone the candidates for an initial screening; however, this task will be assisted in future versions with a telephone button to call the candidate's number. If a candidate passes this screening, the recruiter can schedule an interview; print the resume image, text, or applicant summary; or send an electronic version of the resume text or applicant summary to a hiring manager. All actions are recorded as part of the tracking information stored in an applicant's summary.

Inside Resumix

Inside Resumix are two key technologies: (1) a method for applying both textual and spatial analysis to document understanding and (2) a categorization expert system for categorizing applicant resumes.

Spatial Text Understanding

Resume understanding requires both textual and spatial analysis. Textual analysis is performed by a capable pattern matcher. The pattern matcher in use provides many operators, including a proximity opera-

Experience

1988 - 1990 **Principal Engineer,** XYX Corporation. One of three engineers to design, implement, test, install, and support a law document image retrieval . . .

Figure 2. Experience Section from an Applicant's Resume.

tor useful for context checking; for example, match C only if it appears near another computer language. Patterns can be overloaded with different forms of the same term; for example, match "Relational Database" to "relational database," "RDB," or "RDBMS."

Textual analysis alone does not solve the document-understanding problem. The importance of spatial analysis is demonstrated in the excerpt from a candidate's work history displayed in figure 2.

Clearly, this candidate worked from 1988 to 1990. This information is less obvious to a computer because the starting and ending dates are separated by job-related text. To simplify this problem, Resumix sends text blocks to OCR for character recognition. Blocking information and text are then reassembled into a single document. The translated text fragment from the example in figure 2 is displayed in figure 3.

A pattern matcher uses <BLOCK> patterns to aid extraction. The following pattern might be used as a job indicator:

```
<JOB_INDICATOR> :=
        <BLOCK> <DATE> - <DATE>
```

Blocking information also helps to identify resume sections, for example, objective, experience, education, skills, and references. The following simple-minded pattern might define the experience section:

```
<BEGIN_EXPER> := <BLOCK> EXPERIENCE
<EXPERIENCE_SECTION> :=
        <BEGIN_EXPER> [UP_TO] <END_OF_DOCUMENT>
```

A better job indicator can be constructed:

```
<BETTER_JOB_IND> :=
        <JOB_INDICATOR> [WITHIN]
                <EXPERIENCE_SECTION>
```

The actual patterns used in Resumix can accomodate resumes in almost any format and compensate for common character recognition and text-blocking errors.

The other key technology shipped with Resumix is an expert system capable of categorizing candidate resumes. Resumix currently recognizes candidates in 34 different job categories. Sample categories in-

<BLOCK>
EXPERIENCE
<BLOCK>
1988 -
1990
<BLOCK>
Principal Engineer, XYZ Corporation. One of three engineers to design, implement, test, install, and support a law document image retrieval . . .

Figure 3. Translated Version of Experience Section.

	EXTRACTED DATA	SKILL GROUP	CATEGORY
JOB TITLES:	Accounting Clerk		Finance
	Clerk		Administrative
SKILLS:	Bookkeeping	Accounting	Finance
	Cost Acct	Accounting	Finance
	Debt Leasin	Treasury	Finance
	MVS	IS OS	Info Systems
	Copy	Reprographi	Administrative
	UCC1	Credit	Finance
MAJOR:	General Business		None

Figure 4. Sample of Extracted Data Relevant to Categorization.

clude clerical, information systems, and test engineering. When used in conjunction with skill searches, category searches increase the chance that matching skills will occur in the desired context. For example, a search for the buzzword Fairchild will turn up candidates, spanning 28 categories, who were employed by Fairchild Industries, who live on Fairchild Way, and whose fathers are Mr. Fairchild. If, however, the test engineering category is also required, then responses will be confined to test engineers and technicians. Included will be test engineers trained in the use of Fairchild test equipment, which is what was really desired.

All skills in Resumix are listed in one or more skill groups; for example, “Basic” is listed as a “Programming Language.” Skill groups are listed under one or more job categories; for example, “Programming Language” is listed under “Information Systems” and “Software Engi-

neering." Points are assigned to different categories based on the distribution of skills found in a resume, previous job titles, and college major. A series of rules is applied to produce zero or more job categories. Consider the categorization of a candidate based on the extracted criteria displayed in figure 4.

Job title points are assigned to the "Finance" and "Administrative" categories based on the two previous job titles. Skill points are assigned to the "Finance," "Information Systems," and "Administrative" categories. "Finance" receives additional points because skills occur across multiple financial skill groups. The major "General Business" is not assigned a job category and is ignored. In figure 4, "Finance" receives a strong weighting, "Administrative" receives a weak rating, and "Information Systems" is eliminated as a possible job category. A threshold rule is applied to eliminate "Administrative" as a possible category, and the category "Finance" is assigned. Resumix has a rule base for choosing among combinations of strong and weak categories. The job objective can also affect categorization when a candidate is qualified for two jobs.

The accuracy of this categorization method is a concern for some customers. At AMD, a competition between three senior HR personnel and the Resumix system to classify 40 candidate resumes was deemed a draw, except that Resumix took approximately 20 minutes of clerical time for the entire job, and manual sorting took approximately 1-1/2 hours of each recruiter's time. Former Digital HR Manager Tom Bahlo states that "the key to Resumix's success is its ability to automatically categorize applicants. This task is done by skilled people. Resumix removes these people from the loop."

Criteria for Success

Resumix has performed the tasks for which it was conceived. Recruiters' lives are enhanced because they can concentrate on recruiting instead of paper shuffling. Candidates receive prompt acknowledgments, get a fair reading, and are considered for all positions for which they qualify. Hiring managers benefit from faster response times to open requisitions.

AMD hiring managers have called the HR organization to compliment them on the improved service. One AMD recruiter found matching candidates because a hiring manager described his requirements over the phone. At Sun Microsystems, hiring managers interested in screening their own candidates are being trained to directly use the Resumix system (no recruiters). These benefits alone do not assure a suc-

New Hires
4 hires / month attributed to new data availability
4 / month x 12 x $5K / hire = $240,000

Contract Recruiter
Eliminate 1 recruiter because of improved access
$45 / hour x 2080 = $93,600

Clerical
Eliminate temporary help for college recruiting
1 temp = $10,000

Agency Fees
Reduced agency fees = $25,000

Advertising
Reduce Advertising = $10,000

First-year savings = $378,600

Figure 5. Cost Savings Example for a Mid-Sized System.

cessful application; the following criteria must also be considered: (1) feasibility—system cost versus system savings, (2) extensibility—the ease of expanding to other industries (for example, automotive and medical), (3) maintainability of nationwide installations, (4) connectivity with existing resources, and (5) portability to other platforms.

System cost and savings, the first issue, distinguish between exotic research projects and marketable multimillion dollar products. A cost savings estimate prepared by one Fortune 500 company for its Sunnyvale, California, facility is shown in figure 5. Based on this estimate, the company will recover the cost of the system within a year.

A 2 billion dollar company laid off 30 contract recruiters during a hiring freeze. Only 5 will not be rehired. This number translates to a dollar figure of $2,340,000. Not included are savings in copying and distribution costs (these can exceed $40,000 each year) and the elimination of the need for a separate equal employment opportunity–affirmative action tracking system (Resumix performs this function, saving $10,000 to $15,000).

The second issue is the ability to expand into new industries. Resum-

ix has an ongoing effort to construct a master list of supported job categories. Support for new categories is critical for Resumix to establish its presence across different industries.

The third issue is system maintenance. Resumix requires that all customers provide a dial-in modem line for telephone support. Customers are informed that the line should be disconnected when not in use. Most software problems can be solved over the telephone. For hardware support, Resumix uses a qualified hardware maintenance company to provide hardware support for all equipment used in the Resumix system.

The fourth issue is connectivity. Resumix ties into a company's existing network and passes information through electronic mail. Links to IBM mainframes are supported through TCP/IP-SNA gateways and also over modem lines. Resumix can incorporate customer job and department tables that are dumped to a flat file and can output data in a flat-file format for use in other systems.

The final issue is portability. Some companies such as Digital and Apple Computer refuse to purchase the Sun Microsystems hardware. Internal corporate politics often determine which platforms are acceptable and unacceptable, making system portability an issue. Because Resumix uses its own database and window manager and has source code for interfacing with the OCR unit, porting to the DEC 3100 Unix workstation involved only a few networking and byte-swapping–bit-reversal problems. A port to the IBM personal computer, desirable because of its low hardware cost, proved to be more difficult because of DOS 640K limitations. Ports to the Sequent multiprocessor computer, the Motorola workstation, and the IBM AS/400 are under consideration.

Chronology

Development of the Resumix system began in April 1988. Resumix 2000 was deployed at Sun Microsystems in January 1989 after nine person-months of effort. The system was developed on Resumix's two Sun-3/50 Unix workstations and a third workstation provided by Sun. Resumix 2000 ran on X11r3 and used the Unify relational database. Sun's transition to paperless recruiting occurred slowly over the next two months because of recruiter wariness of the new system and the unavailability of client workstations for all recruiters. For reasons of reliability, efficiency, and portability, X-Windows and Unify were replaced with Resumix software. In May 1989, AMD received the first Resumix 3000 version 1.0 system. Transition to Resumix at AMD occurred within one month. The record for a speedy transition is held by Bank of

America, which stopped its paper shuffling within a week of delivery.

Resumix engineers are currently swamped with hundreds of requests for new and improved capabilities. The company has recently embarked on a major hiring campaign, featuring the slogan "Send us your resume, and we promise not to read it."

Acknowledgments

This system would not be possible without Resumix Inc. founder and president, Steve Leung, who first envisioned this HR aide. Technical leadership and software engineering were provided by Vice Presidents Yul J. Inn and Dave C. Sobotka. Special thanks to Sun Microsystems and AMD for their cooperation and suggestions, which have made Resumix a successful product.