

Online Recruitment - An Intelligent Approach

Samah Rifai and Ramzi A. Haraty
Department of Computer Science and Mathematics
Lebanese American University
Beirut, Lebanon
Email: {samah.rifai, rharaty}@lau.edu.lb}

Abstract - The recruitment process has become extremely tedious, especially with the globalization of markets. The most essential part of the recruitment process, the interview, has not yet been covered in the recruitment automation process. Our “intelligent” system combines both the application form features as well as the interviewing portion of the recruitment process. It also gives suggestions about whom to hire for a particular job vacancy, based on specific and adapted criteria.

Keywords: Intelligence, Hidden Markov Models, and Recruitment System.

I. INTRODUCTION

A. Overview

The job market has changed considerably over the last fifty years. Job seekers are requested to have more specialized skills and a greater amount of knowledge in their fields.

The recruitment process has been severely affected by these changes. Interviewing candidates has become more important than reading their Curriculum Vitae (CV), mainly because the interview enables direct interaction between job seekers and potential recruiters.

With the new advances in technology and communication, oral on-line interviewing, which simulates the actual physical interview, is now possible. Interviewers can now see how candidates interact with their questions and study their body language. However, as noted in [5], not all recruiters are good interviewers, and some interviewers lack the ability to pinpoint keywords that may be pronounced by the candidate and may prove to be essential to the recruitment decision.

There is therefore a need for intelligent analysis of the candidate’s interview answers. This analysis will help the interviewer in the selection process and later on in the final recruitment decision. An intelligent speech recognizer, integrated within the on-line interviewing system, can help achieve this objective.

B. Background

A career can be simply defined as a sequence of employment-related positions, roles, activities and experiences encountered by a person throughout his/her life [1]. Nowadays, a person’s career is seen as more than an occupation or a job, and rather as a life spanning process that covers many events and that undergoes change over time [6].

The Information Technology (IT) revolution had made it absolutely necessary for job seekers to focus on improving their skills and increase their knowledge. Indeed, careers in the twenty-first century are now more oriented towards selecting and managing communications media or knowledge acquisition. Actually, information has become the most important raw material that people need to do their job, and is therefore the primary asset that candidates can use in their job search [6]. This information can be uncovered as early as in the first job interview a candidate can undertake.

When confronted to a career-related decision, individuals are inclined to conduct assessments of their “self”. The “self” reflects one’s personality, which reveals whether one is spontaneous, self confident, free thinking or realistic, one’s values, one’s interests, one’s competencies and finally one’s talents. Candidate selection criteria should therefore be based on a combination of character and personality-related variables.

Having this information in mind, we can design the selection decision making system based on the following variables:

- Interests
- Values
- Abilities/skills/competencies
- Talents
- Objectives
- Self assessment

Interests are activities or subjects to which one feels drawn [1]. They are the easiest indicators for the job

selection process since they can be easily mapped to types of work. Values are guiding principles of life that reveal the individual's priorities in life and influence his/her decisions and actions. Competencies, abilities and skills are closely related. More specifically, competencies represent observable skills or abilities to complete a managerial task successfully. Talents can uncover any manual dexterity, verbal reasoning or interpersonal skills an individual may possess. Self assessment is critical for effective career decisions, since it provides guidelines as well as essential information to potential employers that they are prepared to believe about a job candidate.

Therefore, the candidate's answers shall be categorized into one or more of these categories of variables and grading will be given accordingly. The grading tool we will be using is based on the RIASEC¹ framework proposed by John Holland. It basically suggests six categories of vocational interests that a candidate's interests can be compared against and assessed with. It is based on a wide variety of familiar occupational tasks and day-to-day activities that can be used to produce job titles and prioritize them according to a particular score. Before the interview, the job vacancy is placed into one of the RIASEC categories. During the interview, the emphasis in the candidate's answers should be on the elements related to each category (as depicted in Table 1).

We can generally define an interview as a "meeting of people face to face, as for evaluating a job application" (Collins dictionary, 1981). The core of the interviewing process thus involves managing the information that is given to or delivered by the candidate. Usually this information is extracted by the interviewer and serves as criteria to evaluate the competencies of the candidate with respect to the particular job vacancy. The layout of the interview is presented to the candidate in the form of questions, which may be spontaneous, but in most cases are predetermined before the interview takes place. The conventional interview is usually the central part of the selection process [5]. It is mainly a one-to-one and face-to-face interview with a structured and predetermined set of questions. It can start with a leading question such as "Don't you agree that..." which is then followed by a few closed questions, having either Yes or No as answer, or a straightforward and direct answer. The interviewer then continues by asking some key questions (about ten to fifteen at maximum) [3]. These may be open questions (having lengthy and detailed answers), probing or reflecting questions, in which the candidate is required to dwell on a particular topic such as his/her previous employment experience or academic background.

Table 1. RIASEC Framework (Adapted from [6])

RIASEC category	Types of jobs	Emphasis in resume and interview
Realistic	Manufacturing/Industrial	Tangible results, getting things done
Investigative	Research oriented/Scientific	New ideas, research, scientific evidence
Artistic	Creative and/or Aesthetic	New ways of doing things, aesthetic beauty
Social	Service oriented/Helpful	Benefit to others, personal meaning
Enterprising	Profited oriented/Helpful	Profit and gain, bottom-line results
Conventional	Data/Detail oriented	Accuracy, organization, company policy

Regardless of their order, the interview questions can be divided into several categories. Questions that fall under the critical incident category stress on "incidents that have been significant or critical in determining success or otherwise in the job" [5]. The Job specification and description question category focuses on potential areas of responsibility and organizational structure presented by the candidate or the hiring company. The competencies category of questions focuses on aspects linked to the person's behavior that may meet the job demands within the organization's parameters. The objective of these questions is undoubtedly to determine if the candidate presents essential qualities such as initiative (making decisions and taking charge of events) and influence (ability to convince others to do something by pointing out its benefits).

Based on the above discussion, we will divide the interview questions in our application into several categories, such as critical incident, closed questions, situational, attitudes, intentions, etc... Depending on their relevance and importance for the employment decision, an appropriate grade will be assigned to the interview question.

II. SPEECH RECOGNITION

A. Overview

Speech recognition can be defined as the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program [10]. The first speech recognition applications started in the late eighties, and have gained popularity ever since.

¹ Realistic, Investigative, Artistic, Social, Enterprising, Conventional

Speech recognition applications offer several additional benefits to their end users. First, they offer hands-free computing, which can prove to be extremely useful in environments where keyboards are not practical. Second, they provide users with a more friendly and human-like interface that is more adapted to educational and entertainment applications. Third, they allow easier and faster access to complex lists and choices within forms, since the user simply needs to speak the commands instead of having to select them in repeated steps. Finally, speech-enabled applications can support context sensitive dialogs where the computer's response depends on the user input.

One of the fields where speech-enabled applications have become popular is the medical field, with the replacement of traditional dictation devices with computerized ones that synthesize the spoken input of doctors and transform it into computer-readable format. A typical speech recognition system is designed as in Figure 1.

There are many challenges associated with creating a speech recognition system. First, such system is greatly affected by external conditions such as a noisy environment (*acoustic variability*). Second, variability of speakers can influence the pronunciation and therefore recognition of vocabulary words (*across-speaker variability*). Third, even with the same speaker, variation of tonality, speaking rate and emotional state (*within-speaker variability*) can affect the quality of the acoustic output. Finally, many words in the same language may sound alike (homonyms) but be written in different ways, thus inducing the need for context-specific vocabulary mapping (*phonetic variability*). These factors affect the accuracy of the output and therefore need to be taken into consideration when designing a speech recognition system.

B. Models

There are two main models for representing speech recognition systems. The most famous model is the Hidden Markov Model (HMM) which has been used extensively over the past fifteen years. The second model uses a Neural Network (NN) for the representation. The model to be selected depends mostly on the intent of the application, the size of the vocabulary and the desired performance of the system. NN systems return more accurate matching, while HMM systems are able to handle larger vocabularies. There is therefore a smallest trade off in terms of performance of speech recognition systems. We will adopt an HMM-based system in our application since we will be dealing with a large and diverse vocabulary and the matching obtained is accurate enough to be considered.

HMMs are statistical models that output a sequence of symbols or quantities. These models assume that speech signals could be viewed as piece-wise stationary signals (in time ranges close to 10 milliseconds) that are approximated to stationary processes or states. These states obey the Markov Property, which states that the system “may have changed from the state it was in the moment before, or it may have stayed in the same state” [10]. Also, according to that property, every future state is conditionally independent of every prior state given the current state. Two concurrent stochastic processes are actually defined: the sequence of HMM states (modeling the temporal structure of speech), and a set of state output processes (modeling the [locally] stationary character of the speech signal). The HMM is referred to as a hidden Markov model because “the underlying stochastic process (i.e., the sequence of states) is not directly observable, but still affects the observed sequence of acoustic features”.

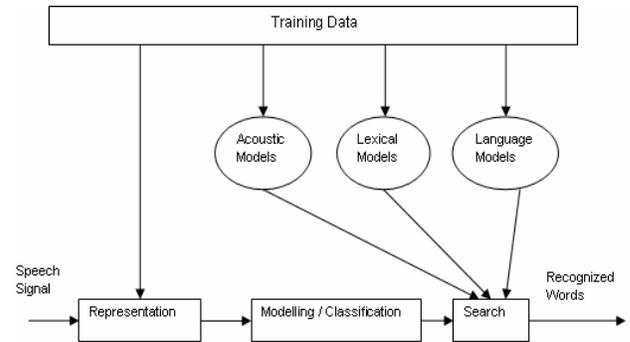


Figure 1. Typical Speech Recognition system

A sentence is typically modeled as a sequence of words. Words in HMM systems are represented by networks of phonemes, corresponding to pronunciations of these words. When a spoken input is pronounced, the acoustic (sound-related) parameters that are attached to that input is seen as a concatenation of a hidden Markov chain (to handle temporal, or time-related variability) and an observable process (to handle spectral, or sound range variability). This allows for managing speech ambiguity and dictionaries having tens of thousands of words.

The HMM system actually computes probabilities or word hypotheses (which can be a single sequence of words, a collection of n-best word sequences, or several overlapping word hypotheses). Generating word hypotheses involves a search process in which a sequence of acoustic features vectors is compared with word models. This means that the probability given by the model to the given sequence will be computed. The detection of word boundaries is itself a probabilistic process that is carried out during the word search.

The HMM model offers a minor disadvantage when the dictionary is more than tens of thousands of words,

since the network becomes too big. One solution is to use multi-pass algorithms, where each pass prepares information for the next, thus reducing the size of the search space. However, in our case, the dictionary is less than ten thousand words since it is mainly restricted within the job and career market domain. This means we can easily use an HMM-based model in our application.

C. Hardware

To be able to capture and interpret sound and speech, computing devices, especially computers, need some essential hardware components.

First, computers need to be equipped with a sound card. Sound cards are dedicated components that synthesize sounds using a technique called Frequency Modulation (FM) synthesis. They also have the capability to record and play digital audio. The quality of a sound card is determined by its resolution, measured in bits (the state of the art is 32 bits). The quality of “human” sound, or sampling, will affect the quality of the speech interface. Sampling rates need to be at least 44 kHz to give the best sound quality. Anything more than 44.1 kHz is deemed unnecessary for human sound (higher rates are reserve for music production).

Second, the most essential hardware component for speech recognition is the microphone. Microphones can control speech user interfaces as well as record human voice. Several types of microphones exist, and different characteristics of these microphones can affect the speech signals. Another important characteristic related to microphones is their physical placement. Some microphones can be built into the computer, while others can be clip-on microphones or attached to headsets. Each application may require a specific microphone type, but the same considerations as for sound cards need to be checked to obtain the best sound quality.

III. USE OF SPEECH RECOGNITION IN THE INTERVIEW PROCESS

A. Technical Considerations

A speech interface is defined as a software interface that employs either human or simulated human speech [7]. A typical representation of a speech interfaces is presented in Figure 2 below.

The speech interface of the recruitment system is associated with the interview module. Our speech enabled web application consists of three components: the ASP.NET server, which holds the application logic and possibly data that are manipulated via scripts, HTML and possibly SALT, the speech server, which recognizes and manipulates speech and audio recordings and finally the

client, which should have the Speech add-in for Internet Explorer installed to be able to run Multimodal speech-enabled web pages. Web pages in the speech-enabled application are developed using an extension of HTML called SALT (Speech Application Language Tags). Microsoft uses the SALT 1.0 specification, member of the World Wide Web Consortium (W3C). W3C is the “main international standards organization for the World Wide Web” [10]. SALT is simply an “extension to HTML that enables developers to add a spoken dialogue interface to Web applications” [7]. SALT represents a subset of XML that adds high-level tags supporting specific elements such as speech recognition input, audio and text-to-speech playback, and dual tone multi-frequency (DTMF) input.

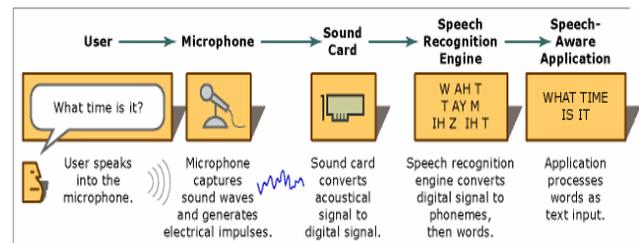


Figure 2. Typical speech interface

The most important type of files included in the speech-enabled application is the grammar. Grammar files have two roles. First, they provide the syntactic rules that “define all possible combination of words and phrases” [7] the user can or is allowed to speak to the application. Second, they add semantic information to the recognized words or phrases.

For the speech applications, grammars are nothing more than XML files that are written in a specific format defined in the World Wide Web Consortium Speech Recognition Grammar Specification Version 1.0 (W3C SRGS). Grammar markups define syntactic rules belonging to the language being recognized.

The grammar in our system obeys the regular English sentence model. In other words, a sentence is made up of a noun phrase followed by a verb phrase. A noun phrase contains zero or one article and adjective followed by a noun. A verb phrase contains a verb, followed by a noun phrase. A sentence can be composite, meaning that it contains a collection of sentences linked by a connector such as “and, or, with, while, if, then, etc....” We have extended the grammar to handle composite sentences, therefore making it as closely related to real conversations as possible.

The details of the grammar are depicted in Figure 3 below.

B. System Design

Ideally, it would become a necessity to have an integrated recruitment system that handles the whole process from the application until the candidate selection. Such an integrated system can be developed as a simple web multimodal application. Multimodal applications have Graphical User Interfaces which guide the user through the application. Users can enter data by speaking to the application (using a decent microphone), typing data into text boxes, or selecting items from drop-down lists or other controls.

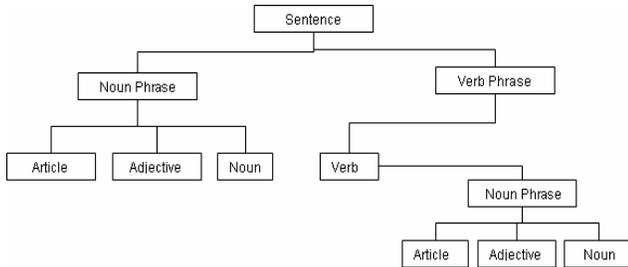


Figure 3. English Grammar

Our integrated recruitment system is built as a modular multimodal application. The first part of the system is the Recruitment module, where the applicant fills an on-line application form containing all the required information. Applications can be filtered in that first stage by an administrator user, to add more restrictions on the candidates that can be accepted for the interview. In the second module, the user may undergo the “intelligent” on-line interview. The system is called “intelligent” because it is automatically responsible for selecting the top three candidates for each job vacancy after analyzing their interviews’ answers and evaluating them appropriately. The evaluation criteria are conveniently tailored to each job vacancy, since selection criteria may vary depending on the job type and required qualifications. Grading is done according to the variables define above.

The Recruitment and Interview modules are designed as simple web applications built using the Microsoft .NET technologies. They communicate via HTTP requests made by the client (the Internet browser) to the server (the physical location where the applications reside).

Once the user has started the interview, a timer initialized to the length of the interview (as defined in the database record for that interview) will be launched. Each question will be displayed on the form and the interviewee can input his/her answer by pressing the “Record your answer button” and then speaking directly into the microphone. To ensure that the interview will be as successful as possible, the candidate should make sure

that the environment where he/she is present is noise-free, that his/her voice intonation is as clear and loud as possible and that the words are pronounced in a distinct fashion, with appropriate spacing between words. When the interviewee has finished recording the current answer, he/she should be click on the “Next question” button to get to and answer the following interview question.

It is worth mentioning that as soon as each interview answer is recorded and saved in the database, the Interview module will calculate the answer’s total grade and add it to its record in the database. Answers and their grades are given special weights (defined prior to the interview) to make sure that most answers will be treated as fairly as possible.

When all interview questions have been recorded and answered, or if the time limit (as specified by the initial timer) has been reached, the interview will be ended automatically and the user will be redirected to a Time Out form that signals the official end of the interview.

The Administrator user will receive the summary of the interview results in the form of a table giving the top three candidates for that interview. This means that the three candidates with the highest scores will be the ones suggested by the system for employment. The administrator can choose to notify these candidates (and those that have not been selected) of the results.

The actual speech recognition occurs during the interview. Whenever the user answers a question, the answer captured will be divided into words. These words will be compared against the grammar structure and then will be graded according to the criteria defined earlier. The accuracy of the recognition is essential since grading will depend on the recognized word.

C. Results

The accuracy of the recognition process which occurs at the interview level is critical to the success of the grading process. The speech recognizer in our case returned an overall accuracy rate close to 80%, which is close to the current accuracy rate (about 85%) of most up-to-date speech recognizers. Smaller words are recognized with a rate close to 95%, while larger and more complex words are recognized with a success rate between 60% and 80%.

We can increase the accuracy level even more by using the pronunciation editor available with the Microsoft Speech Application Development Kit (MS SADK). This gives us the power to train the potential input into distinguishing homonyms (words that sound alike but are written differently) by differentiating their way of pronunciation as well as complex and composite words.

The evaluation and grading component returns the weighted grade of the top three candidates filtered by job vacancy and interview. Based on the system recommendation, the administrator can notify both accepted and not selected candidates, or can choose to perform more detailed manual selection.

IV. CONCLUSION AND FUTURE WORK

The automation of common business processes has revolutionized our world. It has reduced processing time, costs and lowered human-induced errors. Nowadays, with the advent of new technologies, activities that were considered “reserved” for humans, such as conducting interviews, can now be automated, thus saving valuable time and resources for their users.

The system that was presented in this report handles the entire recruitment process, starting by the application form, then the interview and finally until the candidate selection. The interview module, which uses an intelligent speech recognition component, is the core of the Recruitment system. The accuracy rate of the recognition process was close to 80%, which makes the on-line interview process using speech a very efficient tool for the automation of the recruitment process.

We can suggest additional improvements that could be made at a later stage. Ideally, the speech recognition component could be trained to learn new vocabulary items and integrate it into its dictionary. The current trend says that it is better to have a constrained system that works than a fully conversational system that results in errors. Another improvement that could add more value to the system is the ability to generate an ideal candidate profile for each job vacancy. This means that after all candidates have passed the interview for a particular job vacancy, the system could compile all the received answers and suggest ideal answer patterns for each question. These patterns could then later on be fed into the dictionary of Power List items. In this way, the system is learning from previous “experiences”, thus increasing its “intelligence”.

REFERENCES

[1] Arnold, J. *Managing Careers into the 21st Century*. London: Paul Chapman Pub, 1997.

[2] CASCAiD Ltd. ADULT DIRECTIONS - NEW VERSION. Retrieved May 28 from Adult Directions Multimedia: <http://www.careersmatch.co.uk/>, 2006.

[3] Clegg, B. *Instant Interviewing*. London: Kogan Page, 2001.

[4] Cole, R. et.al. *Survey of the State of the Art in Human Language Technology*, 1996, Retrieved May 20, 2007 from <http://cslu.cse.ogi.edu/HLTsurvey/>.

[5] Edenborough, R. *Effective Interviewing: A Handbook of Skills and Techniques* (2nd edition). London: Kogan Page, 2002.

[6] Kummerow, J. *New Directions in Career Planning and the Workplace: Practical Strategies for Career Management Professionals* (2nd edition). Palo Alto CA: Davies Black Pub, 2000.

[7] Microsoft Corporation. *Microsoft Speech Application SDK Help file*, 2005. Retrieved May 20, 2007 from Microsoft Speech Application SDK: www.microsoft.com/speech/Speech/SDK/51/sapi.chm

[8] NY Times, Jobs. Retrieved June 1, 2007, from Post a Job, Find a Job, Get Career Advice - The New York Times & Monster Web site: <http://jobmarket.nytimes.com/pages/jobs/>

[9] RecruitActiv Ltd. *Online Recruitment and Applicant Tracking Systems*. Retrieved June 6, 2007 from web recruitment-online recruitment-applicant tracking system for employers: <http://www.recruitactive.com>

[10] Speech Recognition. In *Wikipedia* [Web]. Wikimedia Foundation inc.. Retrieved May 15, 2007, from http://en.wikipedia.org/wiki/Speech_recognition

[11] Stuart Mitchell and Company Ltd. *Products*. Retrieved May 16, 2006 from Career HORIZONS - career, performance and talent management: <http://www.careerhorizons.net/index.asp>

[12] Top-Career_Resumes.com. *Discover Good Interview Questions You Need to Know About*. Retrieved May 25, 2007 from Good Interview Questions & Answers for The Basic Interview Types: www.Top-Career-Resumes.com,2007.

[13] Weinschenk, S., & Barker, D. *Designing Effective Speech Interfaces*. New York: Wiley, 2000.

ACKNOWLEDGEMENTS: This work was funded by the Lebanese American University.