

PST: The Provider Selection Tool

Howard Marmorstein
Jayesh Ghia
Sandeep Sathaye
Akshay Gupta
Eva Baron-Vartian

Oxford Health Plans
800 Connecticut Avenue
Norwalk, Connecticut 06854
hmarmors@oxhp.com

Abstract

Oxford Health Plans, Inc. is a managed care organization whose goal is to deliver cost-effective, high-quality health care. Oxford's product lines include traditional health maintenance organizations, point-of service plans, third-party administration of employer funded benefit plans, Medicare and Medicaid plans, and dental plans. A Member's satisfaction with their Primary Care Physician (PCP) is important to their relationship with Oxford.

People traditionally choose a doctor (i.e., a PCP) by asking family and friends for recommendations. While this will most likely remain the most trusted method for choosing a Physician, Oxford desires to act as an added supportive resource for helping Members make a wise decision. Oxford aims to complement and also confirm the advice of family and friends. Because of this, Oxford has developed the Provider Selection Tool (PST).

PST is a case based reasoning tool, deployed within Oxford and on the World Wide Web, that elicits search criteria from a Member to assess and evaluate a roster of Providers that meet those criteria. These Providers are then evaluated and presented in ranked order based upon how well they match the criteria. By helping Members easily select the Provider that best meets their criteria, Oxford enhances the one-to-one Oxford-Member relationship.

Problem Description

Why is helping Members find a Physician important?

Both Oxford and independent surveys consistently show that Member satisfaction with a physician is strongly correlated with Member satisfaction with the Health Plan.

This is a difficult situation because Oxford does not have direct control over doctors, yet, doctors have an enormous amount of influence on Oxford's success.

Data collection. Today, the issue is about the power of data. The advantage of having a Physician network is the vast amounts of data amassed on physicians. Oxford has been able to ensure that every Physician in its network delivers only high-quality care; this sets Oxford apart from its competitors. The problem has been helping Members make the right decision when it comes to choosing a new Physician to suit their individual needs. It is no longer good enough to mass market a Physician network; fulfilling the Member's personal and unique needs, especially when it comes to health care, is the key to high satisfaction and keeping the Member at Oxford.

Empowering the health care consumer. It is not enough to give Members a roster, a hard-copy listing of all Oxford participating Physicians, and tell them to choose. In fact, that has been a primary source of dissatisfaction and perpetuates the perception that managed care and forces people to choose doctors who may not be quality providers, despite the claims made. Members looking for a new Physician may, of course, call their Oxford Customer Service Associate (CSA) for more information and guidance, but this also does not solve the problem of creating a good Physician-Member match. The Member still had to pull the names from the roster, then the CSA had to pull the biography for each Physician. The Member received a bit more information about the Physician, but it was still the same process; the same way of thinking.

Now, as Oxford collects more data on Physicians, including clinical statistics and Member satisfaction scores, the old process becomes more cumbersome. It has become more important than ever for the Member to have the ability to describe the characteristics they prefer in a health care provider, and for Oxford to do the work in identifying the Physicians who possess those characteristics. The

answer to the question of "how do we help Members make sense of our data?" is the Provider Selection Tool (PST).

Provider Search Process prior to PST

When a Member asked for assistance in choosing an Oxford participating Physician (i.e., a Provider), a Customer Service Associate (CSA), the Oxford representative who responds to and resolves Member and Provider inquiries across a wide range of subjects, primarily based the decision upon location (e.g., city, zip code, or county). The CSA then looked through a roster to determine the list of Providers that met that location criteria. After which, individual Provider information screens, part of the on-line production system, were invoked to read other relevant detailed information about that Provider. At this point, a dialogue occurred between the CSA and the Member to assess the relevance of that Provider. This process was then repeated until a Provider was identified that sufficiently met the Member's needs. This was a time-consuming and inefficient operation.

Rosters. A roster is divided alphabetically within a county. The CSA must read through all the Providers within a county to find those with the appropriate city and/or zip code. One usually does not spend the time to go through the complete roster before making a decision; this leads to a disproportion amongst Providers chosen by Members based upon last name (i.e., Providers whose last name begins with an *A* may receive more business than those that begin with a *Z*). Physicians whose last name begins with a letter at the end of the alphabet may be at an automatic disadvantage with a hard-copy roster, which creates the potential for high dissatisfaction with the plan among these participating Providers.

Provider Assessment. The assessment across Providers is done solely by the Member; that is, the Member must retain details about a Provider to compare one Provider against another. This makes it difficult to effectively assess all of the requirements.

Application Description

In an effort to strengthen the relationship with Oxford Members, Oxford's Technology Consulting Group has developed a tool for its CSAs. This tool is the Provider Selection Tool (PST).

PST Overview

The Provider Selection Tool assists CSAs to help select the best Provider that meets a Member's criteria. Criteria include: location (city, state, county, zip), Provider name, Provider type (primary care Physician, specialists, dentists, facility, alternative medicine Provider), languages spoken,

years in practice, affiliations, subspecialties, educational background, weekend availability, and Member satisfaction survey results with their Physician.

A CSA elicits information directly from a Member which forms the basis to generate the list of Providers that match. This is done via a case based reasoning (CBR) tool. After which, those Providers are assessed based on matching relevance and, perhaps more importantly, behavior patterns beneficial to the patients via a rule-based system. PST allows one to assign weight to individual criteria, that is, make one attribute more important than another. The model prioritizes the list according to the weight of each attribute.

System Architecture

While designing the system architecture there were three major objectives.

- to provide an effective solution for the needs of the Business Group
- to design a system architecture capable of easy integration with other applications
- to make the application components re-usable

Application Components

PST is divided into three components, each of which are independent of each other.

User Interface. This component allows users to enter the selection criteria for selection of a Provider; it also displays selected Providers.

Intelligent Search Engine. This component is capable of searching Providers intelligently; it is a *Black Box* in that it receives a request from the User Interface, does a search and returns the results back to the requester.

Database. This component stores data relevant to the application.

These components interact with each other by message passing. This messaging architecture makes the application open. Any of the above mentioned components can be replaced without affecting the overall architecture. This approach also makes these components re-usable for other applications.

Tool Selection.

The design of the architecture drives the requirements for selection of the appropriate tools which enable development of the above mentioned components.

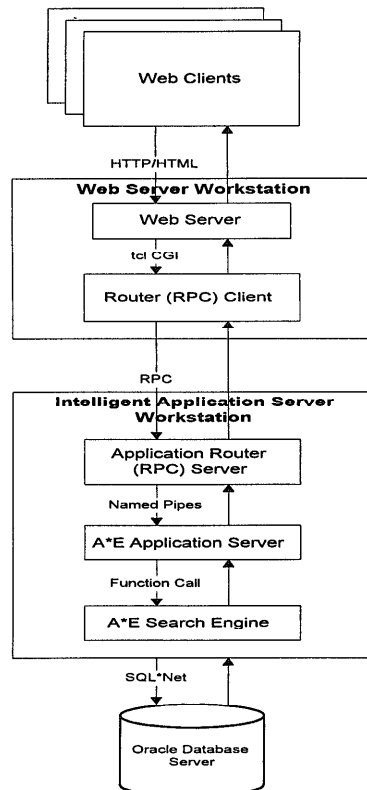
User Interface. The major factor in tool selection was availability. The tool needed to be portable across all platforms, deliverable throughout Oxford, and available for use by Oxford's customer base outside the company. For these reasons a Netscape client was selected.

Intelligent Search Engine. The driving factors for tool selection were the ability to do case based searching with extensibility to rule based reasoning, the ability to rank results depending on the selection criteria, and the ability to do searches quickly. For these reasons the case based reasoning components of ART*Enterprise (A*E) were selected. The advantages of using A*E are many. It provides a development environment which integrates case based reasoning with object oriented, rule based, and procedural programming. Though A*E's Graphical Tool Kit was not used for the final solution, it aided in development of a quick PST prototype during the initial stage of development.

Database. Oxford's production database uses Oracle 7.1. By adhering to the standard makes development consistent across the company.

Application Architecture

There are four distinct components of this architecture:



Web Client. This is the User Interface, developed using tcl and perl scripts, which allows interaction with the Search Engine. It allows one to make a search request and also to examine the search results.

Web Server. A Sun/Solaris machine with Apache serves the Web Clients. This server also runs the Application Router Client program, which directs the user requests received from the Web Client to the Intelligent Application Server. This Application Router Client program also

waits for a response from Application Router. Once it gets a response it sends it to the Web Client. The Application Router Client uses Remote Procedure Calls (RPCs) to communicate between itself and the Application Router.

Intelligent Application Server. This machine runs three processes:

- Application Router, which acts as an interface between the Web Server and the A*E Application Server. This is basically a RPC Server program, written in C, which receives a request from the Application Router Client and passes it to A*E Application Server using named pipes.
- A*E Application Server, which receives the request from the Application Router through named pipes and passes it to the A*E Search Engine. It also passes the results from the A*E Search Engine back to the Application Router through named pipes.
- A*E Search Engine, which uses a CBR paradigm to do an intelligent search on Provider information using a CBR index.

Database Server. An Oracle database server is the source of all the production data required by the A*E Search Engine.

Deployment on the World Wide Web

Upon successful completion of the PST prototype, the challenging task of deploying PST to about 5,000 prospective users within Oxford, along with nearly two million Members outside Oxford, was addressed. The obvious choice was deployment over the World Wide Web (Web).

To achieve this, PST would need to meet several criteria: accept requests over the Web and return the search results as HTML; handle requests from multiple web servers; perform session management as if the server could be multi-threaded; and have an acceptably fast response time. At that time, A*E was not designed to serve documents over the Web. It lacked inter-application communication via sockets or RPC, session management, and multi-threading. In order to enable A*E to function as a server over the Web the following steps were necessary:

- creation of a CGI script (in tcl) that would accept requests over the Web via an HTTP Form
- the CGI script in turn writes the request to a *Request* named pipe
- A*E reads the request from the *Request* named pipe and processes it
- A*E writes the results back out to the *Results* named pipe
- the CGI script reads the results from the *Results* named pipe
- The CGI script generates a results HTML page using the results returned from A*E

Since Oxford has a multiple web server environment, and it was not effective to run the A*E server on each web server, the following scheme was employed:

- code was written in C which acts as a simple client that can send a request to a remote server using RPC; this utility runs on each web server

- the A*E application runs on just one application server
- on that same application, an RPC server utility is run that accepts the requests from the RPC clients called by the CGI scripts on each web server; this RPC server writes the requests to a named pipe *Request*
- the A*E server reads the request and writes the results to the *Results* named pipe

Since PST's initial release, which was written using a pre-production version (2.0 Alpha) of A*E for Solaris, Brightware has added RPC and sockets support, and is also beta testing a new product called ART*Web that will provide for easier deployment of applications on the Web.

Case Based Reasoning

Why a Case Base? PST is a search and assessment tool; not just a lookup and retrieval tool (though it can be used for that purpose). Systems do exist which use traditional methods for data retrieval (e.g., SQL calls to a database). Though extensions for these conventional methods are used (e.g., a Soundex match to account for typographical errors), those tools aren't sufficient.

These tools only retrieve data if the query was successful (i.e., an exact match). However, when a search is done manually, a person can perform a fuzzy match. For example, should a Member wish to find a Neurologist in Darien, CT with 20 years experience who speaks Ukrainian, it is unlikely that an exact match can be made. A person would realize that though exact matches may not be made, there are Neurologists not in Darien, but in the neighboring city of Norwalk. Using a conventional approach no Neurologists would be returned; using a CBR approach those Neurologists that best match the criteria would be returned (e.g., a Norwalk Neurologist with 19 years experience who speaks Russian would be a match; not an exact match, but a partial match). If one would have done a manual search one would eventually relax the requirements to increase the search space. A CBR tool mimics this process.

CBR Advantages. The main advantages of using CBR technology over a simple database query tool are many:

- a CBR tool finds the closest matches rather than exact matches; in other words, unlike a SQL query (which only returns exact matches), it will return a sorted list of Providers in the order of how well they match with the specified criteria.
- a CBR tool allows threshold value definitions, this allows the system to select only those cases whose total weight is above the threshold which provides better control over matches.
- different weights can be assigned to each criterion also allowing better control.
- number matching allows matching within ranges; for example, a match on 10 years of experience, will

return Providers with 10 years of experience followed by those with 9 and 11 years of experience, etc.

- text matching allows a string match (i.e., an exact match), a word-by-word match, or a character-based match which accounts for typographical errors.

CBR Implementation. The CBR component of A*E has been used to build the case base for PST. A*E creates an index on loaded cases which stores each case with specified attributes and their associated weights. This index enables a quick search through the case base to find the matching case. Once this index is created, it can be stored in a file. After which, one only needs to load the index file instead of loading actual cases which would otherwise need much more memory.

In earlier phases of PST development, there was only one index file to store all the Providers. But as the Provider base grew, it started taking up more memory and search time. At this point, it was necessary to breakdown the case base into some logical form. This breakdown was done on the basis of Provider type (e.g., Internal Medicine, Pediatrics, etc.) and networks (e.g., Freedom, Liberty, etc.). For example, the case base index file PEDI-LBITY-CASE-BASE.CBR has only the pediatricians who participate in the Liberty plan. Thus, when a user wants to find pediatricians who participate in the Liberty plan, PST will load only this case base, search through it to find matching Providers which match other user-specified criteria, and then unload the case base index.

Creating Case Bases. The process of creating case bases is a two-stage operation. First, several Oracle stored procedures are executed, which retrieve all the Providers along with the necessary details from different tables. This data is stored in a denormalized table. Second, PST retrieves data from this table and builds the necessary case base files. This process is performed weekly; should the Provider data change too dynamically this process can be implemented nightly.

As mentioned earlier, a separate case base is created for each Provider type and plan type (i.e., network). The system retrieves all the Provider types and plan types from separate Oracle tables along with the conditions needed to be considered while creating that case base. This scheme is robust; one can create new case bases by populating Oracle tables instead of modifying the application code.

Weighting Strategy (i.e., Knowledge)

PST's knowledge lies in the weighting strategy employed and the matches made to the case base. This is the core of the system. The weight strategy is not fixed; it will evolve with the usage of PST. That is, the weighting strategy will be adaptable to become increasingly sophisticated to meet the demands on the system.

All the match criteria along with match type, regular match weight, higher match weight (if selected as important criteria), mismatch weight, minimum precision (for number matching), and maximum deviation (for number matching) are stored in a database table instead of hardcoding it into the program. This makes it easier to change the weights for a particular criteria without making any changes to the code; else every time a weight changes, a new system image would have to be generated.

In the current scheme of weight assignment, location is given the maximum weight; lower weight is assigned to all the other criteria. Currently a zero mismatch weight and a zero absence weight are used for all the criteria. In addition, one can indicate one or more of the criteria to specify as important. Such criteria would then be assigned a higher match weight.

Allowing the Member to state the importance of individual criteria can be realized in different manners. One could use an absolute scale of importance, yet this places too much a burden upon the user. That is, if the user is requested to rank each criterion in order of importance (e.g., city = 1, language = 2, weekend availability = 3, etc.) this would be too time consuming and overbearing. Instead, a simple weighting strategy of important versus non-important was employed. This strategy, though sufficient, is being extended to multiple levels of importance. That is, the user will be able to assign levels of importance to individual criterion (e.g., necessary, important, suggested, or not-interested).

Should the Member not specify many preferences, the chance for multiple Providers having the same overall matching score increases. To alleviate this problem, PST will employ a back-end system. This system will use quality based information (e.g., positive clinical outcomes, member satisfaction scores, etc.) to resolve equality (i.e., ties amongst Providers) when a Member's criteria are equally met.

Location. Character matching (based on trigrams) is used for City and County in order to allow for misspellings. String matching is used for State as it is faster and also it prevents from assigning partial weights (e.g., *CA* should not match *CT*, even partially). Zip code is divided into separate matching criteria: entire zip code (a string match), first three digits of the zip code (a string match), and the last two digits of the zip code (a number match).

Languages, hospital affiliations, qualifications, subspecialties. Character based matching is used for these criteria to allow for typing mistakes. The user can specify one or more values for each criterion. When more than one value is specified for a criterion, the system always *ANDs* these values which may not be the user's intent. This is a limitation of the CBR engine in A*E; that is, A*E does not support *ORing*. For example, a search for a

Provider who is affiliated either with Norwalk hospital OR Stamford hospital, yet the system treats it as Norwalk hospital AND Stamford hospital. This inappropriately gives lower weight to the Providers who are affiliated with only one of the hospitals as compared to the Providers who are affiliated with both hospitals.

Years in Practice. Number matching is used for this criteria. One can specify minimum years and/or maximum years in practice. Since number matching in A*E does not support the concept of minimum and maximum, a work-around is required. An example illustrates the necessity:

If a user put in 6 years as the minimum years of experience, the system should give full match weight to all the Providers who have more than 6 years of experience and full mismatch weight to Providers who have less than 6 years of experience. But the way number matching works, it will give full weight to Providers with 6 years of experience, less weight to Providers with 5 and 7 years of experience, etc. This is not acceptable as Providers with 5 and 4 years of experience should not be selected at all.

As a work-around, the mean of the minimum and maximum years in practice is calculated. To account for empty values, default values, 0 and 30 respectively, have been specified. For example:

A minimum value of 6 and no maximum value are specified. The system would calculate the mean as:
$$\text{mean} = (6+30)/2=18$$

Therefore, Providers with 18 years of experience would be given an exact match, those with 17 and 19 years would be given less weight, etc.

This work-around may not give the exact results but it will prevent undesired results.

Weekend Availability. String matching is used for this criteria. PST is only checking the existence of a value for this field which indicates the Provider has office hours during the weekend.

Member Satisfaction Survey. . Number matching is used for this criteria. Currently there are six criteria specified from a Member Satisfaction Survey:

- Overall satisfaction with Physician
- Likely to recommend Physician
- Physician is skilled and experienced
- Physician fully explained diagnosis/treatment
- Physician shows Member respect
- Physician is accessible in an emergency

The Member can include any of these criteria while searching for a Provider. Assigning a weight to each criteria produced undesirable results. To understand this, consider two different scenarios:

- The user specifies a Location (weight=50) and all of the Member Survey criteria as important (weight=10

*6=60); the impact of the Member Survey exceeds the impact of the Location

- The user specifies a Location (weight=50) and only one of the Member Survey criteria (weight=5); the Member Survey hardly has any impact

In order to use Member Survey effectively, PST assigns a fixed combined weight to Member Survey as an aggregate, rather than assigning individual weights to each Member Survey criterion. The aggregate weight is divided by the number of criteria which are selected to determine the weight for each individual criterion. In addition, those criteria which are indicated as important are given twice the emphasis, and hence, count as twice the number of criteria. This scheme assigns double the calculated individual weight to an important criteria. For example:

The user selects four criteria, one being indicated as important. Assuming an aggregate weight of 15, e.g., PST will assign a weight of 3 ($15/(3+2)$) to three of the criteria while a weight of 6 ($3*2$) is assigned to the criteria indicated as important.

Customized weighting. Plans are underway to extend the weighting strategy from an important/not-important scheme to a multiple level scheme. This scheme would allow the Member to assign levels of importance to individual criterion (e.g., necessary, important, suggested, or not-interested). This would require a change to the internal weighting mechanism. These weights would be stored as part of the Member's profile for future system use.

How Does PST Work?

Upon PST invocation, one is presented with a menu of choices for five categories of participating Providers: Primary Care Physicians and Ob/Gyns, Specialists, Dentists and Dental Specialists, Hospitals and Facilities, and Alternative Medicine Providers.

Provider Search Screen. After the user selects a Provider category, the Provider Search Screen unique to that Provider category is presented:

The current version of PST supports many different search criteria: Provider type, plan type (i.e., network), Provider id, Provider name, location (zip, city, county, state), languages, hospital and practice affiliations, qualifications, sub-specialties, years in practice, and Member Satisfaction Survey results. Not all of these criteria are used as case base attributes; some are for case base selection, some are for lookup, and others are attributes of the case base. For example:

- Provider id. If Provider id is given, a SQL call is made to the database to get the Provider details for the given Provider id. In this case, PST ignores other criteria, even if present.
- Provider type and plan type. These are used to determine the case base file required for loading. For example, if plan type is Liberty (LBTY) and Provider type is pediatrics (PEDI) then the case base required for loading is PEDI-LBTY-CASE-BASE.CBR
- Location, languages, affiliations, qualifications, sub-specialties, years in practice, weekend availability, and Member Satisfaction Survey results. These criteria are used as the actual case base attributes and are assigned weights.

The Provider Search Screen allows one to give extra weight to certain criteria; that is, to make one attribute more important than another. This model prioritizes the returned list according to the weight of each attribute. For instance, if a user indicates that a PCP affiliated with Mount Sinai Hospital is more important than the PCP's ability to speak Spanish, the PCPs affiliated with Mount Sinai will have a higher match weight than a Spanish speaking PCP.

Provider Results Screen. After inputting the criteria and clicking on the search button, the model will search through the loaded Provider case base to find the first 15 Providers which most closely match the profile.

The majority of the development effort was in understanding of the CBR components of A*E and the affect of each case base attribute. Finally, Marketing played a crucial role in helping position PST to the internal customer (i.e., Provider Relations had to signoff on the validity of the knowledge in PST) and to the external customer (i.e., Marketing needed to convey this information to Oxford's Participating Physicians).

Maintenance and Enhancements

Maintenance

PST requires minimal maintenance. PST recreates its case bases weekly. This allows for changes in the Provider tables (e.g., changes in address, creation of new Providers, deletion of old Providers) to be constantly reflected. The general system design is generic; that is, much information and knowledge has been stored externally from the tool in database tables. This scheme allows easy manipulation of the weighting strategy (e.g., if the importance of a search criterion needs to change, one would only have to change a match weight coefficient in a database table). Likewise, if a new Provider specialty needs to be added (as was done with the new roster of Alternative Medicine Providers which Oxford supports), one would only have to create a new case base type entry in a database table. This allows the business areas to easily manipulate and maintain the knowledge without intervention from development staff.

Enhancements

Criteria. The most important of the search criteria is location; all the other criteria are of secondary importance. Therefore, work is being done to enhance its capability. A Geographic Information System (GIS) is being incorporated to perform accurate location matching. A GIS package enables individual addresses, both Member and Provider addresses, to be geocoded (assigned longitude and latitude coordinates); this would allow distance computation between locations. Therefore, the location search could be specified not just as a specific locale, but as a proximity (either in distance or time) to a locale.

Users. Though PST was initially conceived as a tool for Customer Service Associates; it's use has expanded to include other users, both internal to external. Internal users include Medical Management case managers, Behavioral Health case managers, Oxford On-call (Oxford's 24-hour telephone nurse service), Medical Delivery, and Marketing. External users include Members, participating Physicians, and brokers through the Oxford Home Page (<http://www.oxhp.com>) developed by the Electronic Commerce group. Each of these users has criteria specific

to meet their search needs (e.g., Behavioral Health case managers require a more clinical-based search). Plans are underway to identify and meet these needs.

GUI. As Oxford expands its Web Site, the functionality of PST will increase. The GUI will be redesigned (plans are currently underway) to be more robust and *smart* using a language such as Java. This would allow for accurate Member validation and functionality without having to contact Oxford directly (e.g., after the list of Providers appear, the Member could actually select or change their Primary Care Physician without Oxford intervention).

Conclusion

Oxford Health Plans believes in the importance of disseminating information directly to its Members. Use of the World Wide Web is one of the distribution mechanisms. By creating intelligent-based applications for the Web, one can effectively empower the user. The Provider Selection Tool meets this challenge.

Currently, PST is in wide use within Oxford and the number of users via the Web directly by Members is increasing. Reaction to PST has been extremely positive. It provides timely information to Members which enable them to make fully informed decisions in their health care treatment. Initial quantitative results for PST are still forthcoming. The results are being collected, stored in a set of Member profile tables, and will be analyzed for actual effectiveness.

Acknowledgments

The authors want to thank and acknowledge the following individuals who have been involved with PST over the course of the project: development and deployment (Lisa Connelly, Mel Astillero, Steve Kramer, Younes Ghanian, Steven Hoder, Kevin Short, Jonas Rosenthal, Indy Crowley, Don Greist and those at Brightware, Inc.), business knowledge (Roma Virani, Matt Smith, Joe Chrzescijanek, Jim Zdru, Meg Cassidy), sponsorship (Bob Smoler, David Finkel, Michelle Gavens, Steve Daniels, Steve Flannery, Ted Kremer, Katie Krapes), and all the Customer Service Associates without whom this project would never have become a success.

Trademark Acknowledgments

ART*Enterprise is a registered trademark of Brightware, Inc. Oracle and SQL Net are registered trademarks of Oracle Corporation. Solaris is a registered trademark of Sun Microsystems, Inc. Netscape is a trademark of Netscape Communications Corporation.