Abstract
People use arguments to justify their claims. Computer systems use explanations to justify their conclusions. User acceptance of a system depends partly on how well the system explains its reasoning and justifies its conclusions. We have built WOZ, an explanation framework that justifies the conclusions of a clinical decision-support system. WOZ’s central component is the explanation strategy that decides what information justifies a claim. The strategy uses Toulmin’s argument structure to define pieces of explanatory information and to orchestrate their presentation. WOZ uses explicit models that abstract the core aspects of the framework such as the explanation strategy. In this paper, we present the use of arguments, the modeling of explanations, and the explanation process used in WOZ. WOZ exploits the wealth of naturally occurring arguments, and thus can generate convincing medical explanations.

The Proposal – Explanations as Arguments
Clinical practice guidelines encapsulate bodies of medical evidence and other knowledge necessary for decision-making about specific health problems. In the past few years, we have seen considerable interest in decision-support systems that automate clinical practice guidelines. Such systems can be used by clinicians to provide patient care that is informed by the guidelines. These systems especially help inexperienced or non-specialist clinicians to follow advanced evidence-based best practices relevant to their patients. In our laboratory at Stanford Medical Informatics we have developed the EON architecture – a set of software components with which developers can build robust guideline-based decision-support systems [Musen et al. 1996]. Based on EON, we have built ATHENA DSS, a decision-support system for managing primary hypertension at the VA Palo Alto Health Care System [Goldstein et al. 2000]. ATHENA DSS interprets the hypertension guideline model using specific patient data, and provides patient-specific recommendations for clinical management (Figure 1). For example, as shown in Figure 1, ATHENA DSS can recommend adding a specific drug to the patient’s regimen because the patient...
has high blood pressure and diabetes. An important component of ATHENA DSS is the explanation facility.

In conversation, humans use arguments to support beliefs and claims by providing evidences. Lawyers argue their clients’ cases, writers propound their beliefs, and physicians justify their diagnoses. Explanation is the way that a knowledge-based system similarly justifies its conclusions to its users. As requested by the user, the system presents different levels of information in support of its claims. The need for explanations in ATHENA DSS was evident from the survey of 58 primary care health professionals participating in a guideline skills workshop. As part of the survey, they were asked if they would want to see recommendations only or recommendations with explanations, and on the type of explanations they preferred. The majority preferred seeing, at a minimum, the rules used to generate the recommendation. Nearly half wanted the relevant parts of the guideline document to be included in the explanation as well. One third preferred the rules, the guideline document and references to medical literature.

The survey gave us an understanding of what form of explanations would be useful for clinicians using ATHENA DSS. We used it to develop the main design criteria for the explanation function: (1) the explanation should not only provide the rules that lead to the recommendation, but should also offer broader supporting material, such as pointers to the medical literature and the specific guideline document, and (2) the elements of an explanation should be organized and presented so that clinicians could navigate among them effectively. We developed WOZ as a multi-client framework that is part of EON and that generates explanations for decision-support systems based on EON. We adapted the WOZ framework to build the explanation function for ATHENA DSS.

EON’s problem-solving components use explicit models of medical-domain and clinical-protocol knowledge. WOZ, like EON, uses explicit models that abstract the explanation strategy and the visual clients architecture. The explanation strategy defines what information WOZ uses to justify an EON claim. The strategy model uses an argument structure proposed by Toulmin, a philosopher, in his theory of reasoning [Toulmin 1958]. Toulmin’s structure allows the recipient of the argument to identify the different elements needed to support the claim. Central to his logic is his six-element argument structure (Figure 2) by which claims can be argued, regardless of the context of the argument:

1. **Data**: The particular facts about a situation on which a claim is made
2. **Warrant**: The knowledge that justifies a claim made using the data
3. **Backing**: The general body of information or experience that validate the warrant
4. **Qualifier**: The phrase that shows the confidence with which the claim is supported to be true
5. **Rebuttal**: The anomaly that shows the claim not to be true
6. **Claim**: The assertion or conclusion put forward for general acceptance

Wick [Wick 1992] pointed out how early research in explanation has, without stated intent, evolved to engulf Toulmin’s argument structure. Ye [Ye 1995] used Toulmin’s argument structure to study the value of explanation in expert systems for auditing. Ramberg [Ramberg 1996] describes a multiple-explanation construction model that constructs explanations for an expert system in the domain of protein purification. We used Toulmin’s argument structure to identify, organize and present information related to the explanation strategy. In this paper, we explain how WOZ models explanations as arguments and show how the models might be extended to make explanations more persuasive.

**EON, a Knowledge-based System Architecture to Automate Clinical Guidelines**

EON, a knowledge-based system architecture (Figure 3), is a suite of clinical models and software components that can be used to provide physicians with decision-support in guideline-based care. We used the Protégé knowledge-engineering environment [Musen et al. 2000; Noy et al. 2000] to build (1) a patient-data model that describes the structure of data that can be obtained from external sources and used to describe patient situations, (2) a medical-specialty model consisting of taxonomic hierarchies of medical concepts and their relationships, and (3) a guideline model called Dharma that defines the
structure of the guideline knowledge needed to generate recommendations regarding clinical decisions and actions. EON consists of a set of middleware servers that perform the computation necessary to support specific tasks in guideline-based patient care. One such server, the Padda guideline execution server [Tu & Musen 2000], takes as inputs formalized clinical guidelines and relevant patient data to generate situation-specific recommendations. A second server, the Tzolkin temporal data mediator [Nguyen et al. 1999], extends the traditional relational database server to include capabilities to resolve queries involving complex temporal relationships and to create temporal intervals representing abstractions derived from primitive time-stamped data. A third component, WOZ interacts with other components to provide multi-faceted explanations for the recommendations of EON.

ATHENA DSS, a Hypertension Advisory System Based on the EON Architecture

We used the EON components to build the ATHENA DSS application. Using the EON guideline model, we created a computer interpretable representation of major portions of the hypertension guideline described in the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (NIH 1997), and of the VA hypertension guidelines. ATHENA DSS applies relevant patient data to the guideline model, determines if a patient is eligible to be treated under the guideline. It determines the patient’s target blood pressure, decides if his blood pressure is under control, identifies his risk group, and provides drug recommendations and general management messages. It displays the patient-specific treatment information at the point of care, and allows a clinician to interact with the system to modify patient data and fetch updated advisories, or to request explanation on any of the recommendations (Figure 1).

Design Features of the Explanation Function

We wanted our explanation function to cover two broad issues: (1) the location of evidence underlying a recommendation, and its content; (2) the most effective way to display that evidence. With these ideas in mind, we used the following design criteria to develop it:

- The explanation function should identify the types of sources that could provide information for generating explanations.

We wanted ATHENA DSS to generate an explanation that not only shows the rules used to arrive at the recommendation, but also other relevant information. For example, in explaining why the system recommended thiazide diuretic in Figure 1, we wanted to show relevant patient data, including medical evidence that supports using this drug, and links to the appropriate text in the guideline document.
The explanation function should implement a mechanism for specifying what information should be part of an explanation. The type of information that should be included in an explanation depends on the type of recommendation being explained. For example, when explaining a drug recommendation, the explanation function should show drug indications for a patient’s conditions and the evidence for the recommendation, and when explaining a patient’s risk group it should provide a description of the risk group and relevant patient data. Therefore, for each type of recommendation, it needs a template identifying the elements that constitute the explanation.

The explanation function should integrate the different pieces of information coherently. It should use a mechanism for specifying the relationships among the different elements of an explanation. These relationships aid ATHENA DSS in generating a cohesive explanation out of varied pieces of data. For example, by using the relationships among disparate pieces such as patient data, the recommendation and the guideline knowledge, the system can generate a cohesive explanation: Given that the patient has Isolated Systolic Hypertension (ISH), the system recommends adding thiazide diuretic or a calcium channel antagonist because ISH is a compelling indication for thiazide diuretic or calcium channel antagonist.

The explanation function should present the information at an appropriate level of abstraction. Displaying all the pieces of the explanation along with the recommendation could create an information overload. We avoided this problem by clustering the explanation pieces into different levels of abstraction, and presenting them to users one level at a time. This approach allows users to start from a high-level explanation and drill-down to get details. For example, the explanation illustrated in the previous bullet text would be at the top level (also see Figure 1), and the evidence showing that ISH is a compelling indication for thiazide diuretic would be at the next level down.

The explanation function should facilitate recursive explanations. It should chain its explanations together so that some of the pieces of the explanation could, in turn, be explained. For example, a clinician might query why the patient has ISH in the previous examples.

The WOZ Explanation Framework

We built the explanation function for ATHENA DSS based on the WOZ explanation framework (Figure 4). WOZ is broadly made of three layers: the content layer, which constitutes the different sources of information that contribute to the explanation, the presentation layer, which consists of the set of visual clients that display the
information in the content layer, and the strategy layer, which describes the information that should be presented to explain a specific recommendation.

The Content Layer
The information that is used for generating an explanation in ATHENA DSS can be classified as reasoning information, which is used to compute the recommendation and supporting information, which provides evidence for the rules that are used in the computation. The reasoning information includes the patient data, the guideline rules, and the medical domain knowledge (such as drug information). The supporting information includes the guideline document, abstracted peer reviewed summaries of trials, the trial reports, and links to the guideline knowledge model. The reasoning information is provided by the underlying EON components. The guideline-execution component provides reasoning information to support its recommendations. The information includes references to relevant objects such as drug-indication objects in the guideline model, patient data used in the reasoning, appropriate static explanatory messages stored in the guideline model, and dynamically created messages that explain the evaluation of complex criteria. For the support information, we compiled medical literature that provides evidence for the drug recommendations. We also annotated parts of the JNC VI hypertension guideline report, which states medical evidence. We then linked the rules in the guideline model knowledge base to the appropriate evidence literature and to the appropriate parts of the guideline report.

The Presentation Layer
ATHENA DSS uses a set of visual clients that display the explanation information in the content layer graphically. The visual clients can receive messages that contain instructions on what to display. For example, an instruction to the patient-data visual client can be: Display the blood pressures of Joe Smith from June 1999 to May 2000. We built a communication model with a set of terms that will be used in encoding these messages. We built new visual clients such as the one that displays the main recommendations screen (Figure 1). We also created clients by wrapping existing applications with a message-handling layer. For example, Protégé is the knowledge-acquisition tool that we used to create and store the hypertension guideline model. We built a visual client around Protégé to display specific guideline model objects. Another example is utilizing applications such as an internet browser to display medical literature. Each source of information in the content layer is associated with a visual client, and, and we built a clients model that describes the type of information each visual client can display. The communication model and the clients model are stored in the WOZ knowledge base.

The Strategy Layer
The core aspect of our explanation framework is the explanation strategy that defines what constitutes an explanation for a claim. It identifies the different components of an explanation such as the claim itself, the medical evidences that support the claim, the strength of the claim, the evidences that contradict the claim, and the patient data that EON used in determining the claim. This explanation strategy is like that used by people who use arguments to support a claim.

About Arguments
A person states a claim to others, and then uses arguments to increase their belief in the claim. These arguments include presenting evidences related to the claim, generally using a structured format. Arguments are widely used by lawyers and writers. A lawyer first presents a claim in her opening statement at a trial. Then, to prove the claim, she describes the physical evidence that supports the claim. She then identifies the laws apply to the case. She may call experts to support her claim. She may back up the evidence and expert opinions by referring to historical data or cases. Writers sometimes use similar structured arguments when stating their points of view. Physicians may use arguments to present their clinical diagnoses, and may support their conclusions with patient data, medical knowledge, and other related cases. Given this role of arguments in human–human interactions, we can see how arguments can be synonymous with explanations in human–system interactions.

![Figure 5. A meta-argument structure for drug recommendations. An explanation can be generated using the elements of the structure. Its presentation can be organized using the relationships among the elements](image-url)
The Strategy Model

We modeled the explanation strategy as arguments, using Toulmin’s argument structure (Figure 2). The elements of the argument structure for a claim identify the information needed for explanation of that claim. Appropriate WOZ explanation agents provide the needed information, and the explanation strategy can use the relationships among the elements of the argument structure to present the explanation consistently and clearly. We define two types of arguments in our explanation strategy, (1) the meta-argument and (2) the concrete argument.

A meta-argument conceptualizes arguments for a class of claims. We state explicitly the elements of the meta-argument structure with abstract descriptions of appropriate pieces of information in EON’s explanation space. The meta-argument structure for explaining drug recommendations is shown in Figure 5. The meta-argument reads “The drug indication supports the qualified drug recommendation determined using the patient data; the clinical guideline report, the relevant medical literature and reference to guideline knowledge base provide the basis for the drug indication; the claim is not true if the missing data assumption, if any, is inappropriate.” All the argument structures are encoded using the terms in the communication model and are stored in the WOZ knowledge base. Since the explanation strategy is modeled explicitly, we can acquire the meta-arguments from medical experts using a knowledge-acquisition tool.

A concrete argument defines an argument for a specific claim in a class of claims. It follows that a concrete argument is an instance of a meta-argument. We specified a meta-argument for explaining drug recommendations (Figure 5). From this meta-argument, we can derive a concrete argument for explaining a particular drug recommendation such as ‘Add thiazide diuretic’ in Figure 1. The abstract descriptions of the pieces of information in the meta-argument are substituted with the actual information related to the computation of this recommendation. With this concrete argument, WOZ can generate an explanation to support the drug recommendation.

A Dialog

To explain a specific claim, at runtime, WOZ selects the appropriate meta-argument in the strategy knowledge base, identifies the explanation information, obtains the information from appropriate clients, derives the corresponding concrete argument, and generates the explanation by organizing the presentation of the clients (Figure 6). We demonstrate this explanation in a visual dialog between WOZ and a user of EON. The user submits queries and receives explanations using direct manipulation via graphical user interfaces. This dialog is based on the examples that we used previously to illustrate the meta-argument and the concrete argument.

1. EON has determined the patient’s treatment plan, and displays drug recommendations (Figure 1).
2. User is interested in the patient data used by the system to arrive at a specific recommendation ‘Add thiazide diuretic’. User submits the query to the system
3. Wizard receives the details of the query. It recognizes that the recommendation ‘Add thiazide diuretic’ belongs to the class of drug recommendations. It selects the meta-argument of the drug recommendations (Figure 5) from the strategy knowledge base. It then consults the client knowledge base and requests the appropriate clients including EON components to provide the information pieces identified in the meta-argument.
4. Clients fill in the concrete argument for the drug recommendation with realtime values for the ‘Add thiazide diuretic’ recommendation. For example, the

Figure 6. The WOZ Explanation Process. The user submits a query for explanation of a recommendation. The numbered circles refer to the following action points: 1) Wizard selects appropriate meta-argument; 2) Wizard asks appropriate WOZ clients that include EON components to provide information specified in the meta-argument; 3) EON components provide information; 4) Wizard presents the explanation information in a structured manner employing appropriate visual clients.
EON’s database mediator component fills the data element, the Protégé client fills the warrant element, and a document client fills the backing element. The clients also generate visual presentations of these elements.

5. **WOZ** uses the data element of the concrete argument to answer the user. It presents the user the patient data that led to the recommendation i.e. that the drug *thiazide diuretic* was recommended to this patient because the patient has *isolated systolic hypertension, diabetes* and *myocardial infarction*.

6. **User** requests to explain the drug recommendation.

7. **WOZ** taps into the warrant element to answer the user. It displays the Protégé client’s presentation that drug *thiazide diuretic* is compellingly indicated by *isolated systolic hypertension*, and is relatively indicated by *diabetes* and *myocardial infarction*.

8. **User** requests the evidence for this relationship between the drug *thiazide diuretic* and the conditions *isolated systolic hypertension, diabetes* and *myocardial infarction*.

9. **WOZ** directs the user to the information in the backing element. It displays the Document client’s presentation of the clinical guideline document that annotates the relationship. This presentation may also link to any other relevant medical literature.

This dialog demonstrates how the concrete argument structure aids WOZ in deciding what information to present when. A clinician can navigate through this structure to explore the different levels of the explanation (Figure 7).

**Discussion**

We have demonstrated how WOZ can justify claims made by knowledge-based systems such as EON by using arguments. We used Toulmin’s argument structure to integrate and present explanation information from varied sources. Acquiring explanation strategies for particular domains includes identifying the classes of claims that the system need to make and defining one meta-argument for each class. We believe that the number of classes is within reasonable limits for a component-based system such as EON.

We did a pilot study to evaluate our explanation framework. We recruited 20 medical fellows and residents at Stanford University. We presented them recommendation screens without and with explanations. We probed their experience with questions on the effectiveness of the explanation function in improving the credibility of the recommendations, and the types of explanation sources they preferred. We also asked them to rate different aspects of the explanation such as presentation, content, navigation, etc. Majority of the subjects opined that seeing the explanation at different levels of detail built up the confidence in the system’s recommendations. They also liked the idea of zooming into the details of the explanation starting from relevant information encoded in the knowledge base and leading to the actual guideline document and medical literature detailing the evidence behind the recommendation. We believe that the detailed document links will be accessed less by practicing physicians compared to the medical residents.

An evaluation of the explanation function in the actual clinic setting is underway as part of a multi-site evaluation of ATHENA DSS. This evaluation will provide a measure of the usefulness of the explanations in supporting clinicians’ therapeutic decisions. In the rest of the section, we list the different types of evaluation that we are considering.

Access to the explanation function is captured in the program logs, where frequency for each patient and proportion of patients per clinician can be determined. Time spent viewing the explanation function can be measured as well as depth (or links) searched. A
correlation between access to explanatory function and therapeutic decisions can then be examined.

Measures of usefulness of the evidence provided in making a therapeutic decision, in supporting arguments used with patients and in influencing adherence to treatment can be captured by the clinicians response to a specific set of questions with a range of semantically and logically distinct response options [Friedman & Wyatt 1997]. These questions would appear in a box before closing the explanatory function in order to avoid recall bias.

Clinicians’ satisfaction with the explanatory function in terms of ease of access, quality of information, utility of the information, content and visual display can be measured in a survey.

The WOZ explanation process that we presented earlier parallels a dialog that could have taken place between a physician and a patient. Horton [Horton 1998], Dickinson [Dickinson 1998], and Jenicek and Hitchcock [Jenicek & Hitchcock 2004] proposed that physicians use Toulmin’s argument structure to organize medical evidences supporting their diagnoses or treatment plans. These proposals support our contention that it is appropriate to use an argument approach to provide medical explanations.

There are explicit relationships among the elements of Toulmin’s argument structure. We can use these relationships when presenting explanations. Other relationships in the argument structure, however, may have to be defined and considered in the presentation. For example, when there is more than one item in any element of the argument – say the Warrant – in what order do we present these items? How do we expose the relationships among these items? One method is to employ the Rhetorical Structure Theory (RST) [Mann & Thompson 1987] that was developed mainly for text analysis and text generation. RST maintains that, in most coherent discourse, consecutive discourse elements are related by a small set of rhetorical relations that is defined by the theory. Three of these relations are Condition, Elaboration and Sequence. Many natural-language-generation systems rely on the rhetorical relations defined in RST. We can use RST by describing various rhetorical relations among the different items of the warrant. By providing such connectives, we can strengthen the cohesiveness of multiple but related items, thereby enhancing the clarity in the presentation.

There is always a question of how to tailor explanations to the user in a way that will enhance the user’s acceptance of the claims. One approach is to extend the argument structure by defining multiple subarguments. We can envision this structure as multiple argument structures having the same Claim, Data, and Modifier but possibly different Warrant,Backing and Rebuttal. A superimposition of these structures will result in an argument structure that contains multiple subarguments. For the same claim and data, we can then create many subarguments each providing a different flavor of the same argument. When WOZ is providing an explanation to a user, WOZ can employ a suitable subargument, thus providing tailored explanations. Naturally, this design presupposes the existence of a user model that abstracts the user profile and preferences.

Conclusion

We showed how medical explanations could be expressed as medical arguments. Our explanation strategy uses a widely recognized argument structure, and can mirror naturally occurring medical arguments. Our approach is unique in terms of displaying the evidence underlying the system’s recommendations [Schiffmann et al. 1999]. There is a growing research interest in the AI community to build upon the results of the argumentation theory. The notion of argumentation schemes [Reed & Walton 2001] as linguistic forms expressing stereotypical patterns of reasoning is of special interest to us. We can explore different types of argumentation schemes for various contexts of user interaction. These schemes can then be employed to generate more natural arguments when responding to users’ questions. We believe that explicitly displaying evidence empowers clinicians to make decisions based on it, rather than simply on recommendations generated from a practice guideline. The system gives clinicians access to updated evidence, which educates them, and it also offers information that they can use to discuss recommendations with their patients. As a result, the risks of bad outcomes are reduced and adherence to treatment is enhanced.

Acknowledgements

This work has been supported, in part, by grant LM05708 from the National Library of Medicine, and VA grants HSR&D CPG-97-006, CPI 99-275, and RCD-96-301. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References


