A Markup Approach to Surveys and Questionnaires

Jeffrey Hsu
Rutgers University

Murray Turoff
New Jersey Institute of Technology

Survey languages are becoming widely used in the formatting and processing of text, especially with the increased interest in electronic publishing, and the Internet. Descriptive markup is of particular importance, because it allows a description of a document or text element which is independent of its final form and output. The creation of survey questionnaires is an important and widely-used application in the area of text processing, and an important component of any markup language standard should be the inclusion of markups for survey question types. This paper presents the design of descriptive markup functions to support the survey application and to extend the utility of the markup approach. It proposes markups which should serve as an extension to existing markup standards. The benefits of markup command-language methods as opposed to standard direct-manipulation WYSIWYG approaches is also discussed.

Surveys are the cornerstones of behavioral and social science research, as well as an important part of information systems research. They are also widely used for marketing research, such as to obtain opinions and buying preferences about products and services. Surveys are also used to gather data on attitudes, beliefs, preferences, and choices between various options.

Questionnaires are an important method for conducting these surveys. Usually, a survey questionnaire would contain a variety of question types, such as Likert, semantic differential, free response, multiple choice, rank ordering, and others.

The “traditional” method of creating questionnaires is to plan out the questions on paper, and then type them out on a typewriter or word processor. The creation of an entire survey therefore takes a great deal of time if you consider the various steps of designing, typing, formatting, and editing. These paper surveys are then administered manually, by asking subjects to fill in these paper questionnaires, which are then submitted back to the survey administrator for data collection and analysis.

This manual way of creating survey questionnaires is cumbersome and slow. A lot of time is spent on the formatting and presentation of the questions when they are created and typed, instead of focusing only on the content of the questions and the overall objective of the survey. The “traditional” method of survey questionnaire design could be referred to as the “direct manipulation” or WYSIWYG (What You See Is What You Get) approach.

The main objective of this effort was to design a set of markups which provides an easier means for creating survey questionnaires, not only paper survey printouts but also automated online surveys. A command language (markup language) approach is used as opposed to the direct manipulation, WYSIWYG approach. While WYSIWYG ap-

Manuscript originally submitted February 3, 1995; Revised March 17, 1995; Accepted July 22, 1995 for publication.
proaches are generally well suited to novice users, the command language approach often brings about greater efficiency and productivity, especially for more experienced users who possess a good knowledge of an application such as the creation and design of survey questionnaires.

From the viewpoint of practical use, the survey markup language designed here is designed to be easy to use. You need merely to select the question type (element) and then indicate the text and other specifications about the question to be created.

Before proceeding further, it would be useful to define what markups are. Markups are sets of tags, tokens, characters, or specialized commands which are placed into a body of text in order to provide information about the text or other data being processed. A markup could be as simple as a space or line feed, or a complex set of symbols for setting all the formatting details for a text document (Coombs et al., 1987; Goldfarb, 1981; Reinhardt, 1994; Reynolds and DeRose, 1992; Wright, 1992).

Markups enable one to “unlock” the data content of a document. In other words, a document is not just a stream of characters, but rather is a data structure which encompasses a great deal of content (Cronk, 1993; Goldfarb, 1991; Van Herwijnen, 1993; Van Herwijnen, 1990).

The concept of a markup is simple—whenever you write something, you also “mark it up.” Markups can be used to specify the boundaries between words, sentences, and paragraphs, and also indicate the typographical and structural features of a text, such as chapter headings, titles, and indented sections. Some forms of markup allow one to specify the various components of a specific document type, whether it be a book, article, or paper. There are several different kinds of markup which are used.

One commonly used form of markup is punctuation mark. This basically involves adding punctuation to the text, such as commas, periods, question marks, and exclamation points (Coombs et al., 1987).

Presentational markup is another widely used form, which uses formatting commands to enhance the presentation of text. For example, horizontal and vertical spacing, underlining, indenting, and page breaks are added directly into the text to make it more understandable and easy to read. Presentational markup clarifies presentation of the text (Coombs et al., 1987).

Procedural Markup is another form of markup where commands placed in the text indicate how text should be formatted. Basically, it provides instructions to the text formatter, and might include commands to set line spacing, format text, justify (left, right, full) a paragraph, and the like. Usually a certain word processor or formatter has its own set of procedural markup commands (Coombs et al., 1987).

Probably the most important and significant type of markup is descriptive markup. This specifies what text element a unit of data is, and allows you to describe and classify it. A descriptive markup (command) language approach allows the creator of a document to define a number of “element types” or “data structures,” which identifies a text portion as a member of a certain class. For instance, you can specify if a piece of text is a long quotation, table, paragraph, or a footnote (Coombs et al., 1987; Tuck, 1989; Blake, 1989; MacLeod, 1990; Van Herwigen et al., 1990).

The power of markup languages is becoming more apparent for the efficient formatting and processing of text, especially with the increased interest in electronic text manipulation, desktop publishing, the Internet, and the World Wide Web. Descriptive markup is of particular importance, because it allows a description of a document or text element which is independent of its final form and output. One important component of any markup language standard should be the inclusion of markups for survey question types. These descriptive markup functions would be very useful in providing support for survey applications and for extending the usefulness of markup languages.

A practical example of an existing markup language standard is SGML (Standard Generalized Markup Language). This standard for marking up text has been defined by various publications (Association of American Publishers, 1992). A widely used subset of SGML is HTML (HyperText Markup Language), which is used to create hypertext-based documents on the Internet World Wide Web.

Using a descriptive markup language (Coombs et al., 1987) approach to creating surveys appears to be a viable alternative to the manual direct manipulation or WYSIWYG (what you see is what you get) approach which is currently the basis of most text editing systems commonly used for creating surveys.

The descriptive markup (command) language approach allows the creator of a survey to select from a number of “element types” or “data structures,” which classify a text stream as a member of a certain type. A document which has been tagged with descriptive markup can be processed by different kinds of systems, and is independent of its final form and the specific system it is being processed on. Instead of specifying the intricacies of formatting and text presentation, it allows the focus to be entirely on the questions and structure of the questionnaire as a whole. In addition, the markups are independent of the machine or system on which they are being used (Coombs et al., 1987).

One of the most important benefits of a descriptive markup approach to survey design is its ability to minimize cognitive demands. Rather than recalling, selecting, and remembering codes for the creation of entry and coding of procedural markup, there is only one step involved in descriptive markup after recognizing the element type: to use the appropriate markup command together with the text to be “marked up.” This frees the survey creator from formatting concerns inherent in traditional direct manipulation or WYSIWYG methods.

The use of descriptive markup also has the benefits of better maintainability and portability. In terms of maintainability, in the case where the actual formatting or structure of a question (element) type needs to be modified, this can be