

dialog author a much finer client-side control over speech interactions. As used herein, a browser that supports full event and scripting is called an “uplevel browser”. This form of a browser will support all the attributes, properties, methods and events of the extensions. Uplevel browsers are commonly found on devices with greater processing capabilities.

[0052] The extensions can also be supported in a “declarative mode”. As used herein, a browser operating in a declarative mode is called a “downlevel browser” and does not support full eventing and scripting capabilities. Rather, this form of browser will support the declarative aspects of a given extension (i.e. the core element and attributes), but not all the DOM (document object model) object properties, methods and events. This mode employs exclusively declarative syntax, and may further be used in conjunction with declarative multimedia synchronization and coordination mechanisms (synchronized markup language) such as SMIL (Synchronized Multimedia Integration Language) 2.0. Downlevel browsers will typically be found on devices with limited processing capabilities.

[0053] At this point though, a particular mode of entry should be discussed. In particular, use of speech recognition in conjunction with at least a display and, in a further embodiment, a pointing device as well which enables the coordination of multiple modes of input, e.g. to indicate the fields for data entry, is particularly useful. Specifically, in this mode of data entry, the user is generally able to coordinate the actions of the pointing device with the speech input, so for example the user is under control of when to select a field and provide corresponding information relevant to the field. For instance, a credit card submission graphical user interface (GUI) is illustrated in FIG. 5, a user could first decide to enter the credit card number in field 252 and then enter the type of credit card in field 250 followed by the expiration date in field 254. Likewise, the user could return back to field 252 and correct an errant entry, if desired. When combined with speech recognition, an easy and natural form of navigation is provided. As used herein, this form of entry using both a screen display allowing free form actions of the pointing device on the screen, e.g. the selection of fields and recognition is called “multimodal”.

[0054] Referring to FIG. 6, a HTML markup language code is illustrated. The HTML code includes a body portion 270 and a script portion 272. Entry of information in each of the fields 250, 252 and 254 is controlled or executed by code portions 280, 282 and 284, respectively. Referring first to code portion 280, on selection of field 250, for example, by use of stylus 33 of device 30, the event “onClick” is initiated which calls or executes function “talk” in script portion 272. This action activates a grammar used for speech recognition that is associated with the type of data generally expected in field 250. This type of interaction, which involves more than one technique of input (e.g. voice and pen-click/roller) is referred as “multimodal”.

[0055] Referring now back to the grammar, the grammar is a syntactic grammar such as but not limited to a context-free grammar, a N-grammar or a hybrid grammar. (Of course, DTMF grammars, handwriting grammars, gesture grammars and image grammars would be used when corresponding forms of recognition are employed. As used herein, a “grammar” includes information for performing

recognition, and in a further embodiment, information corresponding to expected input to be entered, for example, in a specific field.) A control 290 (herein identified as “reco”) includes various elements, two of which are illustrated, namely a grammar element “grammar” and a “bind” element. Generally, like the code downloaded to a client from web server 202, the grammars can originate at web server 202 and be downloaded to the client and/or forwarded to a remote server for speech processing. The grammars can then be stored locally thereon in a cache. Eventually, the grammars are provided to the recognition server 204 for use in recognition. The grammar element is used to specify grammars, either inline or referenced using an attribute.

[0056] Upon receipt of recognition results from recognition server 204 corresponding to the recognized speech, handwriting, gesture, image, etc., syntax of reco control 290 is provided to receive the corresponding results and associate it with the corresponding field, which can include rendering of the text therein on display 34. In the illustrated embodiment, upon completion of speech recognition with the result sent back to the client, it deactivates the reco object and associates the recognized text with the corresponding field. Portions 282 and 284 operate similarly wherein unique reco objects and grammars are called for each of the fields 252 and 254 and upon receipt of the recognized text is associated with each of the fields 252 and 254. With respect to receipt of the card number field 252, the function “handle” checks the length of the card number with respect to the card type.

#### Generation of Client Side Markups

[0057] As indicated above, server side plug-in module 209 outputs client side markups when a request has been made from the client device 30. In short, the server side plug-in module 209 allows the website, and thus, the application and services provided by the application to be defined or constructed. The instructions in the server side plug-in module 209 are made of a compiled code. The code is run when a web request reaches the web server 202. The server side plug-in module 209 then outputs a new client side markup page that is sent to the client device 30. As is well known, this process is commonly referred to as rendering. The server side plug-in module 209 operates on “controls” that abstract and encapsulate the markup language, and thus, the code of the client side markup page. Such controls that abstract and encapsulate the markup language and operate on the webserver 202 include or are equivalent to “Servlets” or “Server-side plug ins” to name a few.

[0058] As is known, server side plug-in modules of the prior art can generate client side markup for visual rendering and interaction with the client device 30. Three different approaches are provided herein for extending the server side plug-in module 209 to include recognition and audible prompting extensions such as the exemplary client side extensions discussed above. In a first approach illustrated schematically in FIG. 7, the current, visual, server side controls (which include parameters for visual display such as location for rendering, font, foreground color, background color, etc.) are extended to include parameters or attributes for recognition and audibly prompting for related recognition. Using speech recognition and associated audible prompting by way of example, the attributes generally pertain to audible prompting parameters such as whether the