

# US Patent & Trademark Office

## Patent Public Search | Text View

United States Patent Application Publication

20250254377

Kind Code

A1

Publication Date

August 07, 2025

Inventor(s)

SHIN; Yong-wook et al.

### SERVER AND METHOD FOR CONTROLLING SERVER

#### Abstract

A display apparatus and a server which implements an interactive system are disclosed. The server includes a communicator which receives text information corresponding to a user voice collected at the display apparatus from the display apparatus, and a controller which extracts an utterance component from the text information and controls so that a query to search contents is generated using the extracted utterance component and transmitted to an external server which categorizes metadata of the content under each item and stores the same, in which the controller generates the query by adding a preset item to a criteria to search a content, when a number of criteria to categorize the content under an item corresponding to the extracted utterance component is less than a preset number.

**Inventors:** SHIN; Yong-wook (Osan-si, KR), SHIN; Seung-min (Bucheon-si, KR), CHOI; Sung-wook (Seoul, KR), LEE; Hye-jeong (Seoul, KR), CHUNG; Ji-hye (Seoul, KR)

**Applicant:** SAMSUNG ELECTRONICS CO., LTD. (Suwon-si, KR)

**Family ID:** 49918519

**Assignee:** SAMSUNG ELECTRONICS CO., LTD. (Suwon-si, KR)

**Appl. No.:** 19/186426

**Filed:** April 22, 2025

#### Foreign Application Priority Data

KR 10-2013-0001839

Jan. 07, 2013

#### Related U.S. Application Data

parent US continuation 18322300 20230523 parent-grant-document US 12301909 child US 19186426

parent US continuation 17199113 20210311 parent-grant-document US 11700409 child US

18322300

parent US continuation 15174142 20160606 parent-grant-document US 10986391 child US 17199113

parent US continuation 14149083 20140107 parent-grant-document US 9380330 child US 15174142

---

## Publication Classification

**Int. Cl.:** **H04N21/278** (20110101); **G06F3/16** (20060101); **G10L15/08** (20060101); **G10L15/22** (20060101); **G10L15/26** (20060101); **G10L19/26** (20130101); **H04N21/232** (20110101); **H04N21/27** (20110101); **H04N21/422** (20110101); **H04N21/4415** (20110101); **H04N21/482** (20110101)

## U.S. Cl.:

**CPC** **H04N21/278** (20130101); **G06F3/167** (20130101); **G10L15/22** (20130101); **G10L15/26** (20130101); **G10L19/265** (20130101); **H04N21/232** (20130101); **H04N21/27** (20130101); **H04N21/42203** (20130101); **H04N21/4415** (20130101); **H04N21/482** (20130101); **H04N21/4828** (20130101); **G10L2015/088** (20130101); **G10L2015/223** (20130101)

---

## Background/Summary

CROSS-REFERENCE TO RELATED APPLICATIONS [0001] This application is a continuation of U.S. application Ser. No. 18/322,300, filed May 23, 2023, which is a continuation of U.S. Application Ser. No. 17/199,113, filed on Mar. 11, 2021, now U.S. Pat. No. 11,700,409, issued Jul. 11, 2023, which is a continuation of U.S. patent application Ser. No. 15/174,142, filed on Jun. 6, 2016, now U.S. Pat. No. 10,986,391, issued Apr. 20, 2021, which is a continuation of U.S. patent application Ser. No. 14/149,083, filed on Jan. 7, 2014, now U.S. Pat. No. 9,380,330, issued Jun. 28, 2016, and claims priority from Korean Patent Application No. 10-2013-0001839, filed on Jan. 7, 2013, in the Korean Intellectual Property Office, the disclosures of which are incorporated herein by reference in its entirety.

## BACKGROUND

### 1. Field

[0002] Methods and apparatuses consistent with exemplary embodiments a server which constructs an interactive system and a method for controlling the server.

### 2. Description of the Related Art

[0003] With the development of electronic technologies, various types of display apparatuses have been developed and disseminated, and equipped with various functions according to users' demands. Specifically, modern televisions (TVs) provide Internet services by connecting the TV to the Internet, and a user can watch multiple digital broadcasting channels thorough the TV.

[0004] Technology to control display apparatuses according to the voices of users has recently been developed in order to control display apparatuses more conveniently and intuitively. Specifically, the TV may perform functions based on commands such as volume control or channel change issued by the voices of users, by recognizing the user voices.

[0005] However, in the related art, there are limitations where the contents cannot be searched for and provided according to user voices.

## SUMMARY

[0006] Exemplary embodiments overcome the above disadvantages and other disadvantages not described above. Also, the exemplary embodiments are not required to overcome the disadvantages described above, and an exemplary embodiment may not overcome any of the problems described above.

[0007] One or more exemplary embodiments provide a server which searches contents when an interactive system is implemented on the server, and a controlling method thereof.

[0008] According to an aspect of the exemplary embodiment, a server which implements an interactive system with a display apparatus is provided, the server including a communicator which is configured to receive text information corresponding to a user's voice collected at the display apparatus from the display apparatus, and a controller configured to extract an utterance component from the text information and control so that a query to search contents is generated using the extracted utterance component, and transmitted to an external server which is configured to categorize metadata of the content under each item and store the metadata. The controller may generate the query by adding a preset item to a criteria to search a content, when a number of criteria categorizing the content under an item corresponding to the extracted utterance component is less than a preset number.

[0009] The server may additionally include a storage configured to store information regarding the criteria to categorize the contents under each item. The controller may determine whether the number of criteria to categorize the contents under an item corresponding to the extracted utterance component is less than the preset number based on the information regarding the criteria to categorize contents.

[0010] The controller may determine the preset item to be added to the criteria to search contents based on user preference information.

[0011] When the contents search cannot be performed in response to the query generated according to the addition of the preset item to the criteria to search the contents, the controller may regenerate a query to search contents based on the extracted utterance components.

[0012] When an item has less than the preset number of the criteria to categorize contents, the item may be at least one of a broadcasting type item and a screen quality item.

[0013] The preset item may be at least one of broadcasting time item and broadcasting channel item.

[0014] According to an aspect of another exemplary embodiment, there is provided a method for controlling a server which implements an interactive system with a display apparatus is provided. The method may include receiving text information corresponding to a user voice collected at the display apparatus, extracting an utterance component from the text information and generating a query to search a content by using the extracted utterance component, and transmitting the generated query to an external server which categorizes the content under each item and stores the same. The generating may include generating the query by adding a preset item to the criteria to search a content, when a number of criteria to categorize contents under an item corresponding to the extracted utterance component is less than a preset number.

[0015] The generating may include determining whether the number of criteria to categorize the contents under an item corresponding to the extracted utterance component is less than the preset number based on the information regarding the criteria to categorize contents.

[0016] The generating may include determining the preset item to be added to the criteria to search contents based on user preference information.

[0017] When contents search cannot be performed in response to the query generated according to the addition of the preset item to the criteria to search contents, the generating may include regenerating a query to search contents by using the extracted utterance components.

[0018] When an item has less than the preset number of the criteria to categorize contents, the item may be at least one of a broadcasting type item and a screen quality item.

[0019] The preset item may be at least one of broadcasting time item and broadcasting channel

item.  
[0020] According to various exemplary embodiments, a problem of too many search outcomes and subsequently deteriorated efficiency in searching contents are resolved when the contents are searched only based on the utterance components extracted from the user voices. Further, a user of the interactive system can be provided with contents search results that suit his intention when making the utterance, while he can also reduce the frequency of having to interact with the display apparatus. Accordingly, user convenience is improved.

---

## Description

### BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The above and/or other aspects will be more apparent by describing certain exemplary embodiments with reference to the accompanying drawings, in which:  
[0022] FIG. 1 is a diagram of an interactive system according to an exemplary embodiment;  
[0023] FIG. 2 is a block diagram of a display apparatus according to an exemplary embodiment;  
[0024] FIG. 3 is a block diagram of a first server according to an exemplary embodiment;  
[0025] FIG. 4 is a block diagram of a second server according to an exemplary embodiment;  
[0026] FIG. 5 is a diagram illustrating an example of structured metadata according to an exemplary embodiment;  
[0027] FIG. 6 is a detailed block diagram of the second server in FIG. 4;  
[0028] FIG. 7 is a diagram provided to explain an example of a method for adding criteria to search contents according to an exemplary embodiment;  
[0029] FIGS. 8A, 8B, 9A, 9B, 9C and 9D illustrate examples of UI screens provided from a display apparatus according to an exemplary embodiment; and  
[0030] FIG. 10 is a flowchart provided to explain a controlling method of a server according to an exemplary embodiment.

### DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0031] Certain exemplary embodiments will now be described in greater detail with reference to the accompanying drawings.  
[0032] In the following description, same drawing reference numerals are used for the same elements even in different drawings. The matters defined in the description, such as detailed construction and elements, are provided to assist in a comprehensive understanding of the exemplary embodiment. Accordingly, it is apparent that the exemplary embodiments can be carried out without those specifically defined matters. Also, well-known functions or constructions are not described in detail since they would obscure the exemplary embodiment with unnecessary detail.  
[0033] FIG. 1 is a diagram which illustrates an interactive system according to an exemplary embodiment. Referring to FIG. 1, an interactive system **1000** includes a display apparatus **100**, a first server **200**, a second server **300**, and an external server **400**. As used herein, the 'second server **300**' may be referred to as an 'interactive interface device'.  
[0034] The display apparatus **100** may be controlled by a remote controller (not illustrated). Specifically, the display apparatus **100** may perform operations corresponding to remote control signals received from the remote controller (not illustrated). For example, when the display apparatus **100** is implemented as a TV such as the one illustrated in FIG. 1, the display apparatus **100** may perform functions such as power on/off, channel change, and volume control according to remote control signals received from the remote controller (not illustrated).  
[0035] Further, the display apparatus **100** may perform various operations corresponding to user voices.  
[0036] Specifically, the display apparatus **100** may perform functions corresponding to user voices, or may output system replies corresponding to user voices.

[0037] The display apparatus **100** may transmit collected user voices to the first server **200**. When the user voices are received from the display apparatus **100**, the first server **200** converts the received user voices to text information (i.e., texts) and transmits the text information to the display apparatus **100**.

[0038] The display apparatus **100** transmits the text information received from the first server **200** to the second server **300**. When the text information is received from the display apparatus **100**, the second server **300** generates reply data corresponding to the received text information and transmits the reply data to the display apparatus **100**.

[0039] The display apparatus **100** may perform various operations based on the reply data received from the second server **300**. Herein, the reply data may include diverse information related to commands to control the display apparatus **100** to perform specific functions, and system replies outputted from the display apparatus **100**.

[0040] Specifically, the display apparatus **100** may perform functions corresponding to user voices. In other words, the display apparatus **100** may perform functions corresponding to user voices among functions that can be provided from the display apparatus **100**. For example, when user voices indicating a command to change channels are inputted from the display apparatus **100**, the second server **300** may transmit a controlling command to change channels to the display apparatus **100**. Therefore, the display apparatus **100** may change channels based on the received controlling command.

[0041] Further, the display apparatus **100** may output system replies corresponding to user voices. The system replies may be outputted in at least one format of a voice format and a UI screen format.

[0042] For example, when user voices requesting broadcasting program searches are inputted from the display apparatus **100**, the second server **300** may search for contents matching the intention of the user utterance and transmit the search results to the display apparatus **100**. Therefore, the display apparatus **100** may display a UI screen which includes a list of contents searched for in accordance with the intention of the user utterance.

[0043] Thus, the display apparatus **100** may perform various functions corresponding to user voices based on the received reply data from the second server **300**.

[0044] When the intention of the user utterance is related to searching contents, the second server **300** may search contents matching the intention of the user utterance through the external server **400**, and transmit the search results to the display apparatus **100**.

[0045] Herein, the external server **400** may structure metadata and store the structured metadata. For example, the external server **400** may be implemented as a metadata server which structures metadata information regarding an Electronic Program Guide (EPG), music, video on demand (VOD), photographs, and applications, and stores structured metadata.

[0046] Specifically, the external server **400** may divide metadata of contents per item (or field), structure the metadata of the contents based on criteria to categorize the contents under each item, and store the structured metadata. As used herein, the 'item' may be the attribute which categorizes the metadata of contents, and the 'criteria to categorize the contents' may be detailed attributes which divide contents within each field. For example, when the item is 'broadcasting type', the criteria to categorize the contents may be attributes which further divide broadcasting type into first airing broadcasting and re-airing broadcasting. For another example, when the item is a 'screen quality', the criteria to categorize the contents may be attributes which further divide the screen quality into high definition (HD) and standard definition (SD).

[0047] Specifically, the second server **300** may extract utterance components from text information and generate queries to search contents by using the extracted utterance components. Herein, the second server **300** may generate a query by adding preset items as criteria to search the contents and transmit the query to the external server **400**, when the criteria to categorize the contents under an item corresponding to the extracted utterance components are less than a preset number of

criteria.

[0048] For example, it is assumed that a user utters, “Find re-airing programs.” In this case, the second server **300** extracts “re-airing” as an utterance component related to a broadcasting type from the received text information, and determines the criteria to categorize the contents regarding broadcasting type from the stored structured metadata. Because broadcasting type is divided into two criteria such as re-airing and first-airing to categorize the contents, the second server **300** may generate a query to search the contents by adding a broadcasting channel which a user is currently viewing as the criteria to search contents.

[0049] As a result, according to the query received from the second server **300**, the external server **400** may search broadcasting programs provided from a broadcasting channel which a user is currently viewing, and transmit information regarding the searched broadcasting programs (e.g., at least one of title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, broadcasting channel, and thumbnail) to the second server **300**. When search results corresponding to user voices are received from the external server **400**, the second server **300** may transmit reply data so that the display apparatus **100** can output system replies corresponding to user voices to the display apparatus **100**. Accordingly, the display apparatus **100** may output a list of broadcasting programs provided from a broadcasting channel which a user is currently viewing as re-airing programs in response to system replies to a user voice utterance such as, “Find re-airing programs.”

[0050] Thus, the server according to an exemplary embodiment may generate a query by adding preset items as the criteria to search contents, when the criteria to categorize contents under an item corresponding to utterance components extracted from text information are less than the preset number. The above process may solve the following problem. When contents are searched only by utterance components extracted from user voices, efficiency of contents searching may be deteriorated because too many contents are being searched.

[0051] Although FIG. **1** illustrates the display apparatus **100** as implemented as a TV, this is merely exemplary. The display apparatus **100** may be implemented as various electronic devices such as a mobile phone such as a smart phone, a desktop personal computer (PC), a laptop computer, or a Global Positioning System (GPS).

[0052] Further, although FIG. **1** illustrates the first server **200** and the second server **300** implemented as separate servers, this is merely exemplary. The first server **200** and the second server **300** may also be implemented as a single interactive server including both the first server **200** and the second server **300**. In this case, the display apparatus **100** may transmit user voices to the interactive server, and the interactive server may convert user voices received from the display apparatus **100** to text information, and search contents matching intention of the utterance by using the converted text information. When the criteria to categorize the contents under an item corresponding to the extracted intention of the utterance from text information are less than the preset number, the interactive server may generate a query by adding preset items as the criteria to search the contents. Therefore, when the first server **200** and the second server **300** are implemented as one interactive server including both servers, the display apparatus **100** may transmit user voices to the interactive server.

[0053] FIG. **2** is a block diagram of a display apparatus according to an exemplary embodiment. Referring to FIG. **2**, the display apparatus **100** may include an output **110**, a voice collector **120**, a first communicator **130**, a second communicator **135**, a storage **140**, a receiver **150**, a signal processor **160**, a remote control signal receiver **171**, an input **173**, an interface **175** and a controller **180**. Accordingly, the above-noted elements of the display apparatus may take the form of an entirely hardware embodiment such as a processor or circuit(s), an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware.

[0054] Although FIG. **2** illustrates various elements that can be included in the display apparatus

**100**, the display apparatus **100** may not necessarily include all of the elements, and may not be limited to have only the above-mentioned elements. According to the product type of the display apparatus **100**, some elements may be deleted or added, or substituted with other elements.

[0055] The output **110** outputs at least one of voices and images. Specifically, the output **110** may output system replies corresponding to user voices collected in the voice collector **120** in at least one format of a voice format and a UI screen format.

[0056] UI screen may express system replies corresponding to user voices in text format or search results searched by user voices in list format.

[0057] The output **110** may include a display **111** and an audio output **113**.

[0058] Specifically, the display **111** may be implemented as a liquid crystal display (LCD), an organic light emitting display (OLED), or a plasma display panel (PDP), but is not limited thereto.

[0059] The display **111** may provide various display screens that can be provided through the display apparatus **100**. Specifically, the display **111** may create and display UI screens with system replies corresponding to user voices.

[0060] The audio output **113** may be implemented as an output port such as a jack or a speaker, and may output system replies corresponding to user voices in voice format.

[0061] Further, the output **110** may output various contents. Herein, contents may include broadcasting contents and VOD contents. For example, the display **111** may output images constituting contents and the audio output **113** may output audios constituting contents.

[0062] The voice collector **120** collects user voices. For example, the voice collector **120** may be implemented as a microphone to collect user voices, and may be mounted within the display apparatus **100** to constitute an integrated type microphone, or may be implemented to be separated from the display apparatus **100**. When the voice collector **120** is implemented as a separate type, the voice collector **120** may be implemented as a grip type that a user can hold, or a set type that can be placed on a dining-room table or a living-room table. The microphone may be connected with the display apparatus **100** through a wired or a wireless network, and may transmit collected user voices to the display apparatus **100**.

[0063] The voice collector **120** may determine whether the collected voices are user voices, and perform filtering on the collected voices to remove background noises such as air conditioner sounds, vacuum cleaner sounds or music sounds from the collected user voices.

[0064] For example, the voice collector **120** performs sampling on the user voices, and converts the received signals into digital signals when user voices in analog format are inputted. The voice collector **120** calculates the energy of the converted digital signals and determines whether the energy of the digital signals is more than a preset value.

[0065] If the energy of the digital signals is more than a preset value, the voice collector **120** removes noise elements from the digital signals and transmits the filtered signal to the first communicator **130**. At this process, noise elements are sporadic irrelevant sounds that can occur within the home environment, and may include noises such as air conditioner sounds, vacuum cleaner sounds and music sounds. When the energy of the digital signals is less than a preset value, the voice collector **120** waits for another user voice input without processing the digital signals.

[0066] Accordingly, unnecessary electric consumption can be prevented because whole audio processing is not activated when background noises are not included with the user voices.

[0067] The first communicator **130** performs communication with the first server **200** of FIG. 1. Specifically, the first communicator **130** transmits user voices to the first server **200** and receives text information corresponding to user voices from the first server **200**.

[0068] The second communicator **135** performs communication with the second server **300** of FIG. 1. Specifically, the second communicator **135** may transmit received text information to the second server **300**, and receive reply data corresponding to the text information from the second server **300**.

[0069] The first communicator **130** and the second communicator **135** may perform

communication with the first server **200** and the second server **300** by using various communication methods.

[0070] For example, the first communicator **130** and the second communicator **135** may perform communication with the first server **200** and the second server **300** by using wire/wireless local area network (LAN), wide area network (WAN), Ethernet, Bluetooth, Zigbee, universal serial bus (USB), IEEE 1394, and Wifi. Thus, the first communicator **130** and the second communicator **135** may include chips or input ports corresponding to each communication method. For example, when performing communication with a wired LAN method, the first communicator **130** and the second communicator **135** may include wired LAN cards (not illustrated) and input ports (not illustrated).

[0071] The above exemplary embodiment discloses that the display apparatus **100** may mount separate communicators **130**, **135** in order to communicate with the first server **200** and the second server **300**. However, this is merely exemplary. The display apparatus **100** may also communicate with the first server **200** and the second server **300** through a single communication module.

[0072] Further, the above exemplary embodiment discloses that the first communicator **130** and the second communicator **135** communicate with the first server **200** and the second server **300**. However, this is merely exemplary. The first communicator **130** and the second communicator **135** may perform web browsing by connecting with a web server (not illustrated).

[0073] The storage **140** is a storage medium in which various programs necessary for operating the display apparatus **100** are stored, and may be implemented as a non-transitory computer recording medium such as a compact disk (CD), a digital versatile disk (DVD, and hard disk drive (HDD). For example, the storage **140** may include a read only memory (ROM) which stores programs to operate the controller **180** and a random access memory (RAM) which temporarily stores data according to operation of the controller **180**. Further, the storage **140** may include electrically erasable and programmable ROM (EEPROM) which stores various relevant data.

[0074] The receiver **150** receives various contents. Herein, contents may include broadcasting contents and VOD contents, but is not limited thereto.

[0075] Specifically, the receiver **150** may receive contents from broadcasting stations which transmit broadcasting programs by using a broadcasting network or web servers which transmit content files by using the Internet.

[0076] If contents are received from broadcasting stations, the receiver **150** may be implemented as a system which includes a tuner (not illustrated), a demodulator (not illustrated), and a lighter (not illustrated). When contents are received from sources such as web servers, the receiver **150** may be implemented as a network interface card (not illustrated).

[0077] Thus, the receiver **150** may be implemented as various types of receivers according to the exemplary embodiments.

[0078] The signal processor **160** may perform signal processing on the contents so that contents received through the receiver **150** or the interface **175** can be outputted by the output **110**.

[0079] Specifically, the signal processor **160** may perform operations such as decoding, scaling and frame rate converting on the images included in the contents, and may convert the contents to a format that can be outputted in the display **111**. Further, the signal processor **160** may perform signal processing such as decoding on the audio signals included in the contents and convert the audio signals to a format that can be outputted in the audio output **113**.

[0080] The remote control signal receiver **171** receives remote control signals from external remote controllers. The controller **180** may perform various operations based on the remote control signals received in the remote control signal receiver **171**. For example, the controller **180** may perform functions such as power on/off, channel change and volume control according to controlling signals received through the remote control signal receiver **171**.

[0081] The input **173** receives various user commands. The controller **180** may perform operations corresponding to user commands inputted in the input **173**. For example, the controller **180** may perform power on/off, channel change and volume control according to user commands inputted



from the input **173**.

[0082] The input **173** may be implemented as an input panel. The input panel may be configured as a touch pad, a key pad which arranges various function keys, number keys, special keys and character keys, or a touch screen.

[0083] The interface **175** communicates with external devices (not illustrated). External devices (not illustrated) may be implemented as various types of electronic devices, and transmit contents to the display apparatus **100**. For example, when the display apparatus **100** is implemented as a TV, external devices (not illustrated) may be implemented as various types of electronic devices that can connect to the display apparatus **100** and perform functions such as cellular phone, set-top box, DVD player, Blu-ray disk player and gaming machines.

[0084] For example, the interface **175** may communicate with external devices (not illustrated) according to various wired communication methods such as HDMI and USB or wireless communication methods such as Bluetooth and Zigbee. The interface **175** may include chips or input ports corresponding to each communication method. For example, when communicating with external devices (not illustrated) according to the HDMI communication method, the interface **175** may include HDMI ports.

[0085] The controller **180** controls general operations of the display apparatus **100**. The controller **180** may control operations of the output **110**, the voice collector **120**, the first communicator **130**, the second communicator **135**, the storage **140**, the signal processor **160**, the remote control signal receiver **171**, the input **173**, and the interface **175**. The controller **180** may further include ROM and RAM which store programs and data to control the central processing unit (CPU) and the display apparatus **100**.

[0086] Specifically, the controller **180** may control the first communicator **130** to collect user voices input through the voice collector **120**, to transmit the collected user voices to the first server **200**, and receive text information from the first server **200**. The controller **180** may control the second communicator **135** to transmit the received text information to the second server **300** when text information corresponding to user voices is received.

[0087] When reply data corresponding to user voices is received from the second server **300**, the controller **180** may perform various operations based on the received reply data.

[0088] Specifically, the controller **180** may perform functions corresponding to user voices based on the received reply data.

[0089] Reply data may include controlling commands to control the display apparatus **100**. Herein, controlling commands may include commands to perform functions corresponding to user voices among functions that can be operated in the display apparatus **100**. Therefore, the controller **180** may control each element of the display apparatus **100** so that functions corresponding to user voices can be operated based on the received controlling commands from the second server **300**.

[0090] For example, when the display apparatus **100** is implemented as a TV, and the display apparatus collects a user voice saying, “to channel # (#: channel number),” the second server **300** may determine that the intention of the utterance included in the user voice “to channel # (#: channel number)”, to be a request to change channels to channel # (#: channel number), and transmit controlling commands to change channels to channel # (#: channel number) to the display apparatus **100**. Therefore, the controller **180** may control so that channel # (#: channel number) is tuned based on the received controlling commands and the contents provided through the selected channel is outputted.

[0091] However, this is merely exemplary the controller **180** may control each element of the display apparatus **100** so that diverse operations such as power on/off or volume control can be performed according to collected user voices.

[0092] Further, reply data may include diverse information related with system replies which are outputted in the display apparatus **100**. In this case, reply data may further include controlling commands to output system replies corresponding user voices.

[0093] Specifically, when the utterance of the user voices have the intent to search contents are collected in the display apparatus **100**, the second server **300** may determine the intent of the user utterance, and search contents matching the determined intention of the utterance. The second server **300** may transmit at least one of information regarding searched contents such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, broadcasting channel and thumbnail.

[0094] In this case, the controller **180** may output system replies corresponding to user voices in at least one of a voice format and a UI screen format by using received information. According to the above exemplary embodiment, the controller **180** may generate a list of contents by using information regarding contents received from the second server **300** and output a UI screen including the generated list. The controller **180** may add at least one of information regarding the searched contents such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, broadcasting channel and thumbnail to the list.

[0095] The reply data may include system reply data which expresses system replies outputted on the display apparatus **100** in text format. Therefore, the controller **180** may control the output **110** to output system replies corresponding to user voices in at least one format of a voice format and a UI screen format by using received reply data.

[0096] For example, the controller **180** may convert system reply data expressed in text format to voices by using a text to speech (TTS) engine, and output the voice utterance through the audio output **113**. The TTS engine is a module which converts texts to voices. The TTS engine may convert texts to voices by using various TTS algorithms disclosed in the related art. Further, the controller **180** may create a UI screen which includes texts constituting system reply data, and output the texts through the display **111**.

[0097] The controller **180** may analyze user preference information and store the user preference information in the storage **140**. Specifically, when a command to power on or change channels is inputted, the controller **180** may determine contents provided from the selected channel based on EPG information, and store at least one of information regarding the determined contents such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel. Further, when a command to power on or change input ports is inputted, the controller **180** may determine that contents provided from external devices (not illustrated) connected to the interface **175**, and store at least one of information regarding the determined contents such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel. The controller **180** may request transmission information regarding contents stored in the external devices (not illustrated).

[0098] The controller **180** may determine the title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, and broadcasting channel with regard to contents which are viewed more than a preset number of times to be the title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel which a user prefers, and store that information in the storage **140**. When a command to request user preference information is received from the second server **300**, the controller **180** may transmit user preference information to the second server **300**.

[0099] However, the above is merely exemplary. The controller **180** may display a UI screen which receives user preference information according to preset events, and store the user preference information inputted through the UI screen in the storage **140**.

[0100] The controller **180** may transmit information regarding a channel which a user is currently viewing to the second server **300**. For example, when a requesting command is received from the second server **300**, the controller **180** may determine the currently selected channel through the receiver **150** and transmit information regarding the determined channel to the second server **300**.

[0101] FIG. 3 is a block diagram of the first server according to an exemplary embodiment.

Referring to FIG. 3, the first server **200** includes a communicator **210** and a controller **220**.

[0102] The communicator **210** performs communication with the display apparatus **100**. Specifically, the communicator **210** may receive user voices from the display apparatus **100** and transmit text information corresponding to the received user voices to the display apparatus **100**. The communicator **210** may include diverse communication modules.

[0103] The controller **220** controls general operations of the first server **200**. Specifically, when user voices are received from the display apparatus **100**, the controller **220** controls the communicator **210** to generate text information corresponding to the received user voices, and transmit the text information to the display apparatus **100**.

[0104] Specifically, the controller **220** may generate text information corresponding to the received user voices by using a speech to text (STT) engine. Herein, the STT engine is a module which converts voice signals to texts. The STT engine may convert voice signals to texts by using various STT algorithms disclosed in the related art.

[0105] For example, the controller **220** determines voice intervals by examining the start and end of voices which a user utters within the received user voices. Specifically, the controller **220** may calculate the energy of the received voice signals, classify the energy level of the voice signals according to the calculated energy, and examine the voice interval through dynamic programming. Further, the controller **220** may generate phoneme data by examining the phoneme as minimum unit of voices based on an acoustic model within the examined voice interval, and convert user voices to texts by applying a Hidden Markov Model (HMM) probabilistic model to the generated phoneme data.

[0106] FIG. **4** is a block diagram of the second server according to an exemplary embodiment. Referring to FIG. **4**, the second server **300** includes a communicator **310** and a controller **320**.

[0107] The communicator **310** performs communication with the display apparatus **100**. Specifically, the communicator **310** may receive text information corresponding to user voices collected in the display apparatus **100** from the display apparatus **100**. The communicator **310** may transmit reply data corresponding to the received text information to the display apparatus **100**.

[0108] Further, the communicator **310** performs communication with the external server **400** of FIG. **1**. Specifically, the communicator **310** may transmit a query to search contents to the external server **400**, and receive the search results according to the query from the external server **400**.

[0109] The communicator **310** may include various communication modules. For example, the communicator **310** may perform communication with the display apparatus **100** and the external server **400** through communication methods such as wire/wireless LAN, Ethernet, Bluetooth, Zigbee, USB, IEEE 1394, and Wifi, but not limited thereto. The communicator **310** may mount chips or input ports corresponding to each communication method. For example, when performing communication with wired LAN method, the communicator **310** may arrange wired LAN cards (not illustrated) and input ports (not illustrated).

[0110] However, this is merely exemplary. The communicator **310** may also install other separate communication modules to perform communication with each of the display apparatus **100** and the external server **400**.

[0111] The controller **320** controls general operations of the second server **300**. Specifically, the controller **320** may extract utterance components from text information, generate query to search contents by using the extracted utterance components, and transmit the query to the external server **400**. The external server **400** may divide and store metadata regarding contents per item.

[0112] Specifically, when text information is received from the display apparatus **100**, the controller **320** may extract utterance components from the text information and determine the intention of the user's utterance by using the extracted utterance components.

[0113] For example, it is assumed that text information such as "Find re-airing programs" is received from the display apparatus **100**. In this case, the controller **320** may extract the term, "re-airing programs," as the utterance component related with a broadcasting type from the text information written as "Find re-airing programs" by using prestored keywords. The controller **320**

may determine that the term, “Find,” corresponds to a search request using prestored keywords, and determine that the intention of the utterance “Find re-airing programs” corresponds to a search request for re-airing programs.

[0114] If the intention of the user utterance relates to searching contents, the controller **320** may generate a query to search contents by using the extracted intention of the utterance. The controller **320** determines items which are mapped with the extracted utterance components and criteria which divide contents per item in structured metadata stored in the external server **400**.

[0115] Based on criteria to categorize the contents under an item mapped with the extracted intention of the utterance, the controller **320** determines the attributes effectiveness of searching the features in the extracted utterance components. Specifically, according to whether the number of criteria to categorize the contents under an item mapped with the extracted utterance components is a preset number or not, i.e., when the number of criteria to categorize contents is less than the preset number, the controller **320** may determine the extracted utterance components to be an un-matched searching condition. The preset number may be two. However, the preset number may not be limited to two.

[0116] If the extracted utterance components are determined to be an un-matched searching condition, the controller **320** may generate a query by adding at least one of semantically orthogonal attributes from the extracted utterance components as criteria to search contents. Herein, semantically orthogonal attributes are attributes which do not affect the intention of the user utterance for searching contents and may be already established.

[0117] Specifically, the controller **320** may generate a query by adding preset items as criteria to search contents, when the number of criteria to categorize contents under an item corresponding to the extracted utterance components is less than the preset number.

[0118] Herein, items whose criteria to categorize contents are less than the preset number may be at least one of broadcasting type item and screen quality item. Further, preset items may be at least one of broadcasting time item and broadcasting channel item.

[0119] In the following, it is assumed that the external server **400** structures and stores metadata as illustrated in FIG. 5 for further specific explanation.

[0120] First, the external server **400** may divide metadata of contents per item. Items are various contents information included in metadata, and the external server **400** may divide and store metadata into items such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, and broadcasting channel, as shown in FIG. 5.

[0121] The external server **400** may structure and store metadata of contents according to criteria to categorize the contents under each item.

[0122] For example, referring to FIG. 5, the external server **400** may divide metadata of contents into title item **510** which categorizes the metadata regarding contents by applying “title” as criteria to categorize contents, cast item **520** which categorizes the metadata regarding the contents by applying “cast” as criteria to categorize contents, producer item **530** which categorizes the metadata regarding contents by applying “producer” as criteria to categorize contents, genre item **540** which categorizes the metadata regarding the contents by applying “genre” as criteria to categorize the contents, broadcasting type item **550** which categorizes the metadata regarding the contents by applying “broadcasting type” as criteria to categorize contents, broadcasting time item **560** which categorizes the metadata regarding contents by applying “broadcasting time” as criteria to categorize the contents, screen quality item **570** which categorizes the metadata regarding contents by applying “screen quality” as criteria to categorize contents, viewing rate item **580** which categorizes the metadata regarding contents by applying “viewing rate” as criteria to categorize the contents, and broadcasting channel item **590** which categorizes the metadata regarding contents by applying “broadcasting channel” as criteria to categorize contents. After dividing, the external server **400** may structure and store the metadata of contents.

[0123] However, this is merely exemplary. The external server **400** may also structure and store

metadata by using various other contents information constituting the metadata.

[0124] When the external server **400** structures and stores the metadata illustrated in FIG. 5, a method for generating query to search the contents in the controller **320** will be explained below.

[0125] For example, when text information written as “Find re-airing programs” is received from the display apparatus **100**, the controller **320** may extract the utterance components, “re-airing programs,” from “Find re-airing programs.”

[0126] In this case, “re-airing programs” is an utterance component which is related to the item “broadcasting type”, and an item mapped with the utterance components in the structured metadata corresponds to the broadcasting type item **550**. Therefore, the controller **320** checks the number of criteria to categorize contents under the broadcasting type item **550** in the structured metadata.

[0127] Because the broadcasting type item **550** is divided into two criteria to categorize contents, i.e., first airing and re-airing, the controller **320** may determine that the number of criteria to categorize contents under an item mapped with the extracted utterance component is less than a preset number, and that the extracted utterance components are un-matched searching conditions.

[0128] The reason why the extracted utterance components are determined to be un-matched searching conditions when the number of criteria to categorize the contents under an item mapped with the extracted utterance components is less than the preset number is that efficiency of searching contents becomes lower when too many contents are to be searched when contents are searched only with the extracted utterance components.

[0129] Therefore, the controller **320** may generate a query for searching contents by adding preset items as criteria to search contents. Specifically, the controller **320** may establish searching conditions under the preset items and generate a query for searching contents by using the established searching conditions and the extracted utterance components.

[0130] Herein, preset items may be at least one of the broadcasting time item and the broadcasting channel item.

[0131] Conditions established per preset item may be previously established in each item. For example, regarding the broadcasting time item, conditions may be preset as contents which are aired within a preset time, e.g., 24 hours, from the current time. Regarding the broadcasting channel item, conditions may be preset as contents which are aired on a channel that a user is currently viewing.

[0132] Therefore, when the broadcasting time item is added as criteria to search contents, the controller **320** may establish contents which are aired within the preset time from the current time as conditions regarding the broadcasting time item. The controller **320** may generate a query for searching contents of which broadcasting time is 24 hours from the current time and broadcasting type is re-airing, and transmit the generated query to the external server **400**.

[0133] The external server **400** may perform the searching of contents based on the query received from the second server **300**, and transmit the search results to the second server **300**. The external server **400** may transmit information regarding the searched contents such as at least one of title, thumbnail, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate, and broadcasting channel as search results. For example, when it is 10 o'clock, 7th of January, the external server **400** may transmit title information regarding the searched contents, i.e., Title\_BBB, Title\_CCC by using the structured metadata in FIG. 5.

[0134] When the broadcasting channel item is added as a criteria to search the contents, the controller **320** may establish information regarding a channel that a user is currently viewing as conditions regarding the broadcasting channel item. The controller **320** may request information regarding the currently selected channel in the display apparatus **100** and receive information regarding the channel that a user is currently viewing. Therefore, when the display apparatus **100** currently selects CH 22, the controller **320** may generate a query to search contents of which the broadcasting channel is CH 22 and broadcasting type is re-airing, and transmit the generated query to the external server **400**.

[0135] The external server **400** may perform searching of the contents by using the query received from the second server **300** and transmit the search results to the second server **300**. The external server **400** may transmit information regarding the searched contents such as at least one of title, thumbnail, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel as search results. For example, the external server **400** may transmit title information of the searched contents, i.e., Title\_III by using the structured metadata in FIG. 5.

[0136] As described above, the reason why the broadcasting channel item or the broadcasting time item is added as additional conditions to search the contents is that the intention of the user utterance to search the contents may not be affected. Thus, since the intention of the user utterance in the above exemplary embodiments is to simply request a search for re-airing programs, the intention of the user utterance may be affected when items such as title, cast, producer and genre are established as additional conditions. Therefore, the broadcasting channel item or the broadcasting time item which has a lower possibility of affecting the intention of the user utterance may be added.

[0137] The controller **320** may generate reply data by using the search results received from the external server **400**, and transmit the generated reply data to the display apparatus **100**.

[0138] Reply data may include diverse information related to system replies outputted in the display apparatus **100**. In this case, reply data may further include controlling commands to output system replies corresponding to user voices.

[0139] According to the above exemplary embodiment, when the display apparatus **100** collects "Find re-airing programs," the controller **320** may transmit information of the received contents such as at least one of title, thumbnail, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel to the display apparatus **100**. In this case, the controller **320** may transmit a controlling command to output a contents list to the display apparatus **100** by using the contents information.

[0140] Therefore, the display apparatus **100** may output system replies corresponding to user voices based on the reply data received from the second server **300**. In other words, the display apparatus **100** may output a UI screen constituting a list of the searched contents in response to a user voice request for searching contents. The display apparatus **100** may add information of the searched contents such as at least one of title, thumbnail, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel to the list.

[0141] According to the above exemplary embodiment, the display apparatus **100** may output a list constituting contents BBB and contents CCC or a list constituting searched contents in response to user voices saying, "Find re-airing programs."

[0142] Therefore, according to an exemplary embodiment, when the extracted intention of the utterance is an un-matched searching conditions, contents may be searched by automatically adding criteria to search the contents.

[0143] Although the above exemplary embodiment discloses that an item mapped with the extracted utterance components is the broadcasting type item, this is merely exemplary. The screen quality item **570** is divided into HD and SD, i.e., two criteria to categorize the contents, when utterance components related to screen quality are extracted, the extracted utterance components may be determined to be un-matched searching conditions.

[0144] Further, although the above exemplary embodiment discloses that one preset item is added as a criteria to search the contents, this is merely exemplary. The controller **320** may add a plurality of preset items as criteria to search contents. Specifically, the controller **320** may establish conditions in each of the broadcasting time item and the broadcasting channel item, and generate a query to search contents by using the plurality of established conditions and the extracted utterance components.

[0145] Further, although the above exemplary embodiment discloses that preset items are added as a criteria to search the contents, this is merely exemplary. The controller **320** may determine preset

items to be added as a criteria to search the contents based on user preference information. Herein, user preference information may include at least one of title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel which a user prefers. In this case, user preference information may be received from the display apparatus **100**.

[0146] For example, it is assumed that genre preferred by a user is movie. The controller **320** may determine the genre item to be added as a criteria to search contents and generate a query to search contents by using the determined criteria to search contents. According to the above exemplary embodiment, the controller **320** may establish “movie” as conditions regarding the genre item, generate a query to search the contents of which the genre is movie and broadcasting type is re-airing, and transmit the generated query to the external server **400**. Accordingly, the external server **400** may transmit title information of the searched contents, Title\_HHH by using the structured metadata of FIG. 5.

[0147] Further, although the above exemplary embodiment discloses that, when the extracted utterance components are un-matched searching conditions, contents may be searched by automatically adding other criteria to the search contents, this is merely exemplary. The controller **320** may ask a user whether to add other criteria to the search contents and determine that other criteria to search the contents are added based on user replies.

[0148] Specifically, when the extracted utterance components are un-matched searching conditions, the controller **320** may generate and transmit reply data to the display apparatus **100** so that the display apparatus **100** can output a question asking the user whether to add criteria to search the contents as system replies.

[0149] The controller **320** may express system replies outputted in the display apparatus **100** in text format and transmit the replies to the display apparatus **100**. For example, the controller **320** may express “Your command is simple and may have many results. Broadcasting time is additionally limited. Add and search?” in text format and transmit the text to the display apparatus **100**.

[0150] Therefore, the display apparatus **100** may output a question asking whether to additionally add criteria to search contents based on the received reply data from the second server **300** as system replies. For example, the display apparatus **100** may output UI screen including the text, “Your command is simple and may have many results. Broadcasting time is additionally limited. Add and search?”

[0151] The controller **320** may perform the searching of the contents by adding criteria to search the contents based on the received text information. For example, when text information of “Yes” is received from the display apparatus, the controller **320** may determine that the term, “Yes,” corresponds to a positive response based on prestored keywords, and may perform the searching of contents by adding the preset item as a new criteria to search contents. When text information of “No” is received, the controller **320** may determine that the term, “No,” corresponds to a negative response based on prestored keywords, and perform the searching of the contents by using the extracted utterance components only without adding new criteria to search the contents.

[0152] Further, although the above exemplary embodiment discloses that other criteria to search the contents are automatically added when the extracted utterance components are un-matched searching conditions, this is merely exemplary. When the contents that are searched according to the extracted utterance components do not provide any results, the controller **320** may perform searching contents of the contents by using other parts of the extracted utterance components.

[0153] For example, it is assumed that text information of “Find re-airing programs at CH 7 which all ages can watch” is received from the display apparatus **100**. The controller **320** extracts the term, “CH 7,” as utterance components related with the broadcasting channel, extracts the term, “all ages” as utterance components related with the viewing rate, and extracts the term, “re-airing programs,” as utterance components related with the broadcasting type from the text information of “Find re-airing programs at CH 7 which all ages can watch.” Therefore, the controller **320** may determine that the intention of the utterance of “Find re-airing programs at CH 7 which all ages can

watch” is to request the search for re-airing programs on CH 7 that all ages can watch.

[0154] In this case, the controller **320** may generate a query to search contents by using the extracted utterance components. According to the above exemplary embodiment, the controller **320** may generate a query to search the contents of which the broadcasting channel is CH 7, the viewing rate is all ages, and the broadcasting type is re-airing, and transmit the generated query to the external server **400**.

[0155] However, when the external server **400** arranges the structured metadata of FIG. 5, because there are no contents satisfying the query received from the second server **300**, the external server **400** may transmit a reply to the second server **300** indicating that there are no search results.

[0156] The controller **320** may regenerate a query to search the contents by using other parts of the extracted utterance components. Specifically, the controller **320** may remove the extracted utterance components corresponding to at least one of the broadcasting time type and the broadcasting channel type, and regenerate a query to search the contents by using the other parts of the utterance components. According to the above exemplary embodiment, the controller **320** may remove the utterance components corresponding to at least one of the broadcasting channel, the viewing rate, and the broadcasting type, and regenerate a query to search the contents by using the other parts of the utterance components. In other words, the controller **320** may generate a query to search the contents of which viewing rate is all ages and broadcasting type is re-airing after removing the utterance components regarding the broadcasting channel, or generate a query to search the contents of which the broadcasting channel is CH 7 and the broadcasting type is re-airing after removing the utterance components regarding the viewing rate, or generate a query to search the contents of which the broadcasting channel is CH 7 and the viewing rate is all ages after removing the utterance components regarding the broadcasting type.

[0157] When searching of the contents is not performed with the query generated by adding preset items as criteria to search contents, the controller **320** may regenerate the query to search the contents by using the extracted utterance components. When there are no contents which satisfy the extracted utterance components, and all of the searching conditions have been added as criteria to search contents, the controller **320** may regenerate the query to search the contents by using only the extracted utterance components.

[0158] According to the above exemplary embodiment, it is assumed that a channel which a user is currently viewing is CH 2. In this case, the controller **320** may generate a query to search the contents of which the broadcasting channel is CH 2 and the broadcasting type is re-airing, and transmit the generated query to the external server **400**.

[0159] However, when the external server **400** arranges the structured metadata of FIG. 5, the external server **400** may transmit replies saying that there are no search results to the second server **300** because there are no contents which satisfy the received query from the second server **300**. In this case, the controller **320** may regenerate a query to search the contents by using the extracted utterance components only. Thus, according to the above exemplary embodiment, the controller **320** may generate a query to search the contents of which broadcasting type is re-airing and transmit the generated query to the external server **400**.

[0160] FIG. 6 is a detailed block diagram of the second server illustrated in FIG. 4. Referring to FIG. 6, the second server **300** may further include a storage **330** as well as the elements illustrated in FIG. 4. The like elements of FIG. 6 overlapping with those FIG. 4 with the same function will not be explained for the sake of brevity.

[0161] The storage **330** may store information regarding criteria to categorize contents under an item. The storage **330** may store each item of the structured metadata stored in the external server **400** and information regarding criteria to categorize the contents which further divide each item. Therefore, the controller **320** may determine whether criteria to categorize the contents under an item corresponding to extracted utterance components are less than a preset number based on information regarding criteria to categorize the contents.



[0162] Further, the storage **330** may store diverse information to determine the intention of the user utterance by using received text information from the display apparatus **100**.

[0163] Specifically, the storage **330** may store various information to analyze domain, user action and slot from user voices by using received text information from the display apparatus **100**.

[0164] Herein, the domain may be divided by topics which user uttering voices belong to such as “broadcasting” and “device control.” User action indicates the intention of the user utterance such as “information output” and “device control,” and slot indicates information which specifies object information regarding the domain, i.e., the intention of the user utterance applied within the domain.

[0165] Specifically, the storage **330** may analyze user action within the domain and store keywords to extract the slot.

[0166] For example, the storage **330** may store information which keywords such as “limit,” “search,” “find,” and “show” within the domain correspond to a user action which requests information searching. The storage **330** may store various keywords related with contents such as title, cast, producer, genre, broadcasting type, broadcasting time, screen quality, viewing rate and broadcasting channel regarding contents, and store information with regard to the slot that each keyword corresponds to. According to an exemplary embodiment, the storage **330** may store information where the term, “re-airing program,” is a keyword related to the a broadcasting type which corresponds to a slot.

[0167] According to another exemplary embodiment, the storage **330** may store information where keywords such as “turn on,” “high,” “low,” “turn off,” and “execute” within the device control domain correspond to a user action to request device control. The storage **330** may store various keywords related with device control such as channel name, channel number, volume, and power and store information where each keyword corresponds to a slot.

[0168] According to another exemplary embodiment, the storage **330** may store information where the terms, “Yes” and “Okay,” correspond to keywords which have a positive meaning, and the terms, “No” and “Don’t,” correspond to keywords which have a negative meaning.

[0169] Further, the storage **330** may arrange a corpus database. Herein, a corpus database may be implemented in a format which stores exemplary phrases and replies thereof.

[0170] Thus, the storage **330** may store a plurality of exemplary phrases and replies to phrases per purpose domain. In this case, the storage **330** may tag and store information to analyze exemplary phrases and replies corresponding to the exemplary phrases per exemplary phrase.

[0171] For example, the storage **330** may store an exemplary phrase, “Find re-airing programs,” in a broadcasting domain. In this case, the storage **330** may tag and store information to analyze the exemplary phrase, “Find re-airing programs,” with a corresponding phrase. Specifically, the storage **330** may tag and store a corresponding phrase with information where “re-airing programs” indicates a slot related to a broadcasting type and “Find” indicates a slot related to a request searching contents from the exemplary phrase, “Find re-airing programs.”

[0172] According to another exemplary embodiment, the storage **330** may store an exemplary phrase, “Turn on CH O,” in a device control domain. In this case, the storage **330** may tag and store a corresponding phrase with information to analyze the exemplary phrase “Turn on CH O.” Specifically, the storage **330** may tag and store a corresponding phrase with information where “CH O” indicates a slot related to a channel number and “Turn on” indicates a slot related to a request to power on/off the television from the exemplary phrase, “Turn on CH O.”

[0173] Therefore, the controller **320** may extract utterance components from text information received from the display apparatus **100** by using information stored in the storage **330**, and determine the intention of the user utterance by using the extracted utterance components.

[0174] For example, it is assumed that the text information “Find re-airing programs” is received from the display apparatus **100**. The controller **320** may extract “re-airing programs” as utterance components from the phrase “Find re-airing programs” by considering that the term, “re-airing

programs,” is a slot related to broadcasting type. The controller **320** may determine that “Find” corresponds to a request to search information in a user action, and that the intention of the utterance “Find re-airing programs” is to request the searching of re-airing programs. The controller **320** may determine that the search terms related to the user voice are un-matched terms, generate a query to search contents by adding new searching conditions, and transmit the generated query to the external server **400**. The controller **320** may further transmit reply data to the display apparatus **100** so that the display apparatus **100** can output system replies corresponding to “Find re-airing programs.”

[0175] For another example, it is assumed that text information of “Turn on CH O” is received from the display apparatus **100**. The controller **320** may extract “CH O” as utterance components from “Turn on CH O” by considering that the term, “CH O,” is slot related with channel number. The controller **320** may determine that “Turn on” corresponds to changing channels in a user action and that the intention of the utterance “Turn on CH O” is to request the changing of channels to CH O. The controller **320** may transmit a controlling command for the display apparatus **100** to change channels into CH O to the display apparatus **100**.

[0176] FIG. 7 is a diagram of a method for adding criteria to search contents according to an exemplary embodiment.

[0177] Since the method for adding criteria to search contents is described by referring to FIGS. 4 to 6, it will be briefly explained by referring to FIG. 7.

[0178] The second server **300** extracts utterance components from user voices at operation S720 when text information corresponding to user voices is received at operation S710. The second server **300** determines attributes effectiveness of the extracted utterance components at operation S730. Specifically, the second server **300** may determine attributes effectiveness of the extracted utterance components based on whether criteria to categorize the contents under an item corresponding to the extracted utterance components are less than the preset number.

[0179] Therefore, when the extracted utterance components are determined to be matching searching conditions at operation S730-Y, the second server **300** may generate a query to search the contents by using the extracted utterance components, and transmit the generated query to information sources, i.e., the external server **400**.

[0180] When the extracted utterance components are determined to be un-matched searching conditions at operation S730-N, the server **300** may determine items to be added as criteria to search contents at operation S740. At this step, the second server **300** may generate a question asking a user whether to add criteria to search contents at operation S750, and provide the generated question to a user through the display apparatus **100**. Based on the user's reply to the question, the second server **300** may add criteria to search contents to the extracted utterance components, or generate a query to search contents only with the extracted utterance components, and transmit the generated query to information sources, i.e., the external server **400**.

[0181] FIGS. 8A, 8B, and 9A-9D illustrate examples of UI screens provided from the display apparatus according to an embodiment.

[0182] For example, referring to FIG. 8A, it is assumed that a user utters the phrase “Find re-airing programs.”

[0183] In this case, the display apparatus **100** may display a UI screen **810** including a list of contents based on reply data received from the second server **300** referring to FIG. 8B. The list of contents may show contents searched for by adding at least one of the broadcasting time and the broadcasting channel as criteria to search contents.

[0184] Referring to FIG. 9A, it is assumed that a user utters, “Find re-airing programs.”

[0185] In this case, if the search condition is improper, the display apparatus **100** may display a UI screen **910** asking whether to add criteria to search contents based on reply data received from the second server **300** referring to FIG. 9B. Referring to FIG. 9C, when a user utters, “Yes,” the display apparatus **100** may display UI screen **920** including a list of contents based on reply data received

from the second server **300**. In this case, the list of contents may show contents searched by adding the broadcasting time as criteria to search contents.

[0186] FIG. **10** is a flowchart which illustrates a method for controlling the server according to an exemplary embodiment.

[0187] At operation **S1010**, text information corresponding to user voices collected in the display apparatus is received from the display apparatus.

[0188] At operation **S1020**, utterance components are extracted from the text information and a query to search contents is generated by using the extracted utterance components.

[0189] Specifically, when criteria to categorize the contents under an item corresponding to the extracted utterance components are less than a preset number, a query is generated by adding preset items as criteria to search contents.

[0190] Herein, items of which criteria to categorize contents are less than the preset number may be at least one of the broadcasting type item and the screen quality item. Further, preset items may be at least one of the broadcasting time item and the broadcasting channel item.

[0191] The server may determine whether criteria to categorize contents under an item corresponding to the extracted utterance components based on preset information regarding criteria to categorize contents are less than the preset number.

[0192] Further, the server may determine the preset items added as criteria to search contents based on user preference information.

[0193] Meanwhile, when searching contents is not performed with query generated by adding preset items as criteria to search contents, query to search contents may be regenerated by using the extracted utterance components.

[0194] At **S1030**, the generated query is transmitted to the external server which categorizes metadata of the content under each item and stores the same.

[0195] The method for generating a query by considering preset items or user preference information is described above by referring to FIGS. **1** to **9**.

[0196] Further, the method may provide non-transitory computer readable medium which store programs implementing the controlling method according to the exemplary embodiment.

[0197] Non-transitory readable medium indicate medium which store data semi-permanently and can be read by devices, not medium storing data temporarily such as register, cache, or memory. Specifically, the above various applications or programs may be stored and provided in non-transitory readable medium such as CD, DVD, hard disk, Blu-ray disk, USB, memory card, or ROM.

[0198] Further, the above block diagrams illustrating the display apparatus and the server do not describe a bus; however, communication between the elements in the display apparatus and the server may be performed through the bus. Each device may further include processors such as CPU and micro processor which perform the above various steps.

[0199] Further, the foregoing exemplary embodiments are merely exemplary and are not to be construed as limiting. The present teaching can be readily applied to other types of apparatuses. Also, the description of the exemplary embodiments is intended to be illustrative, and not to limit the scope of the claims.

## Claims

**1.** A display device comprising: a signal receiver; a communicator; a voice input receiver; a display; and a processor configured to: based on a first user voice input corresponding to a control of a channel received through the voice input receiver, change a channel, and control the display to output content corresponding to a broadcasting signal of the changed channel received through the signal receiver; based on a second user voice input corresponding to the content being received through the voice input receiver while the content is being output, receive data corresponding to the

second user voice input from a server through the communicator; and control the display to output a UI included in an item corresponding to a plurality of pieces of content together with the content that is being output based on the received data, wherein the received data is generated based on the second user voice input and additional information.

**2.** The display device as claimed in claim 1, wherein the additional information includes information about a channel in which the content that is being output to the display is broadcast.

**3.** The display device as claimed in claim 2, wherein the information about the channel is transmitted to the server through the communicator.

**4.** The display device as claimed in claim 2, wherein the additional information further includes information about a current time.

**5.** The display device as claimed in claim 1, wherein the UI included in the item corresponding to the plurality of pieces of content includes at least one of titles of the plurality of pieces of content, thumbnails corresponding to the plurality of pieces of content, channels in which the plurality of pieces of content are broadcast, or times on which the plurality of pieces of content are broadcast.

**6.** The display device as claimed in claim 1, wherein the plurality of pieces of content included in the UI correspond to a plurality of pieces of broadcast content that are broadcast in the changed channel within a preset time from a current time.

**7.** The display device as claimed in claim 1, wherein the content that is being output to the display includes content that is received through the communicator, not the signal receiver, and output to the display.

**8.** The display device as claimed in claim 1, wherein the received data is automatically generated based on the additional information together with the second user voice input without receiving any further user input.

**9.** The display device as claimed in claim 1, wherein the processor is further configured to, based on a third user voice input corresponding to a control of a function, not the content, being received while the content is being output, receive data corresponding to the third user voice input from the server through the communicator, and perform the function corresponding to the third user voice input based on the received data.

**10.** The display device as claimed in claim 9, wherein whether data received from the server is generated based on the additional information is based on whether a user voice input received through the voice input receiver corresponds to the content or corresponds to the control of the function.

**11.** A method for controlling a display device including a voice input receiver, a signal receiver, a display and a communicator, the method comprising: based on a first user voice input corresponding to a control of a channel received through the voice input receiver, changing a channel, and outputting, through the display, content corresponding to a broadcasting signal of the changed channel received through the signal receiver; based on a second user voice input corresponding to the content being received through the voice input receiver while the content is being output, receiving data corresponding to the second user voice input from a server through the communicator; and outputting, through the display, a UI included in an item corresponding to a plurality of pieces of content together with the content that is being output based on the received data, wherein the received data is generated based on the second user voice input and additional information.

**12.** The method as claimed in claim 11, wherein the additional information includes information about a channel in which the content that is being output to the display is broadcast.

**13.** The method as claimed in claim 12, wherein the information about the channel is transmitted to the server through the communicator.

**14.** The method as claimed in claim 12, wherein the additional information further includes information about a current time.

**15.** The method as claimed in claim 11, wherein the UI included in the item corresponding to the

plurality of pieces of content includes at least one of titles of the plurality of pieces of content, thumbnails corresponding to the plurality of pieces of content, channels in which the plurality of pieces of content are broadcast, or times on which the plurality of pieces of content are broadcast.

**16.** The method as claimed in claim 11, wherein the plurality of pieces of content included in the UI correspond to a plurality of pieces of broadcast content that are broadcast in the changed channel within a preset time from a current time.

**17.** The method as claimed in claim 11, wherein the content that is being output to the display includes content that is received through the communicator, not the signal receiver, and output to the display.

**18.** The method as claimed in claim 11, wherein the received data is automatically generated based on the additional information together with the second user voice input without receiving any further user input.

**19.** The method as claimed in claim 11, wherein the method further comprises based on a third user voice input corresponding to a control of a function, not the content, being received while the content is being output, receiving data corresponding to the third user voice input from the server through the communicator, and performing the function corresponding to the third user voice input based on the received data.

**20.** The method as claimed in claim 19, wherein whether data received from the server is generated based on the additional information is based on whether a user voice input received through the voice input receiver corresponds to the content or corresponds to the control of the function.

---