

Technical Disclosure Commons

Defensive Publications Series

January 2022

Conversational Document Processing

Rui Costa

Anthony Okwechime

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Costa, Rui and Okwechime, Anthony, "Conversational Document Processing", Technical Disclosure Commons, (January 11, 2022)

https://www.tdcommons.org/dpubs_series/4843



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Conversational Document Processing

ABSTRACT

Filing documentation typically requires a user to laboriously fill in a form with fields for data that are sourced from and/or proved by other documents. This disclosure describes techniques that leverage conversational artificial AI to guide users through the process of identifying and submitting digital copies of documents for an application. A conversational interface as described herein enables organizations to use a single framework to address complex document processing use cases. The techniques can substantially simplify and expedite paperwork processes.

KEYWORDS

- Conversational artificial intelligence
- Conversational interface
- Conversational platform
- Natural language understanding (NLU)
- Document processing
- Online form

BACKGROUND

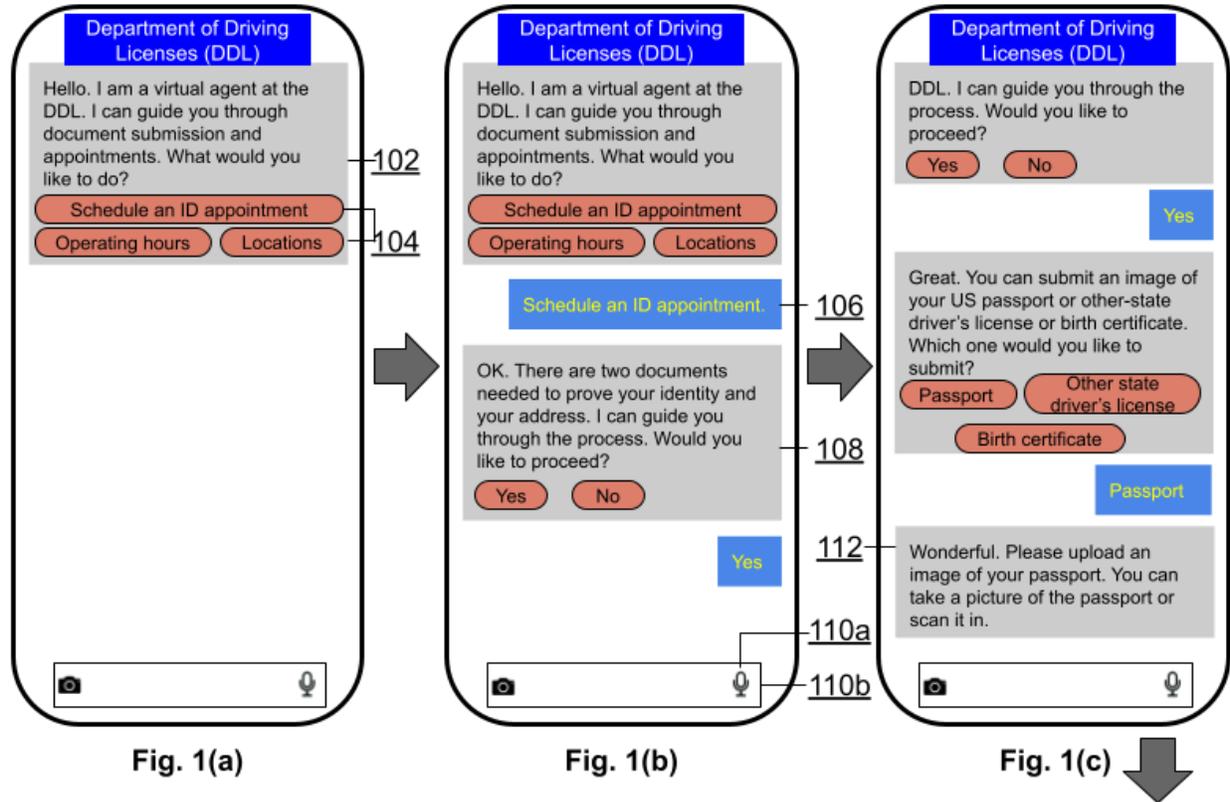
Filing documentation for various reasons , e.g., to pay taxes, to obtain a driver’s license or passport, to apply for a loan, to apply for college, etc., typically requires a user to laboriously fill in fields with data sourced from and/or proved by other documents (e.g., date of birth proved by a government-issued birth certificate, social security number proved by a card issued by social security administration, address proved by a utility bill, income proved by a tax form,

etc.). Today, organizations are left to develop multiple disparate solutions to address complex use cases that involve processing such documentation.

Conversational artificial intelligence (AI) enables human-like conversation to assist users with their input queries. Conversational AI in the form of chatbots or virtual agents is used by organizations to assist customers. These input queries can be simple utterances or more complex use cases.

DESCRIPTION

This disclosure describes techniques that leverage conversational AI to guide users through the process of identifying and submitting digital copies of documents for an application. The techniques bridge the gap between conversational AI and document processing to enable organizations to use a single framework to address complex document processing use cases.



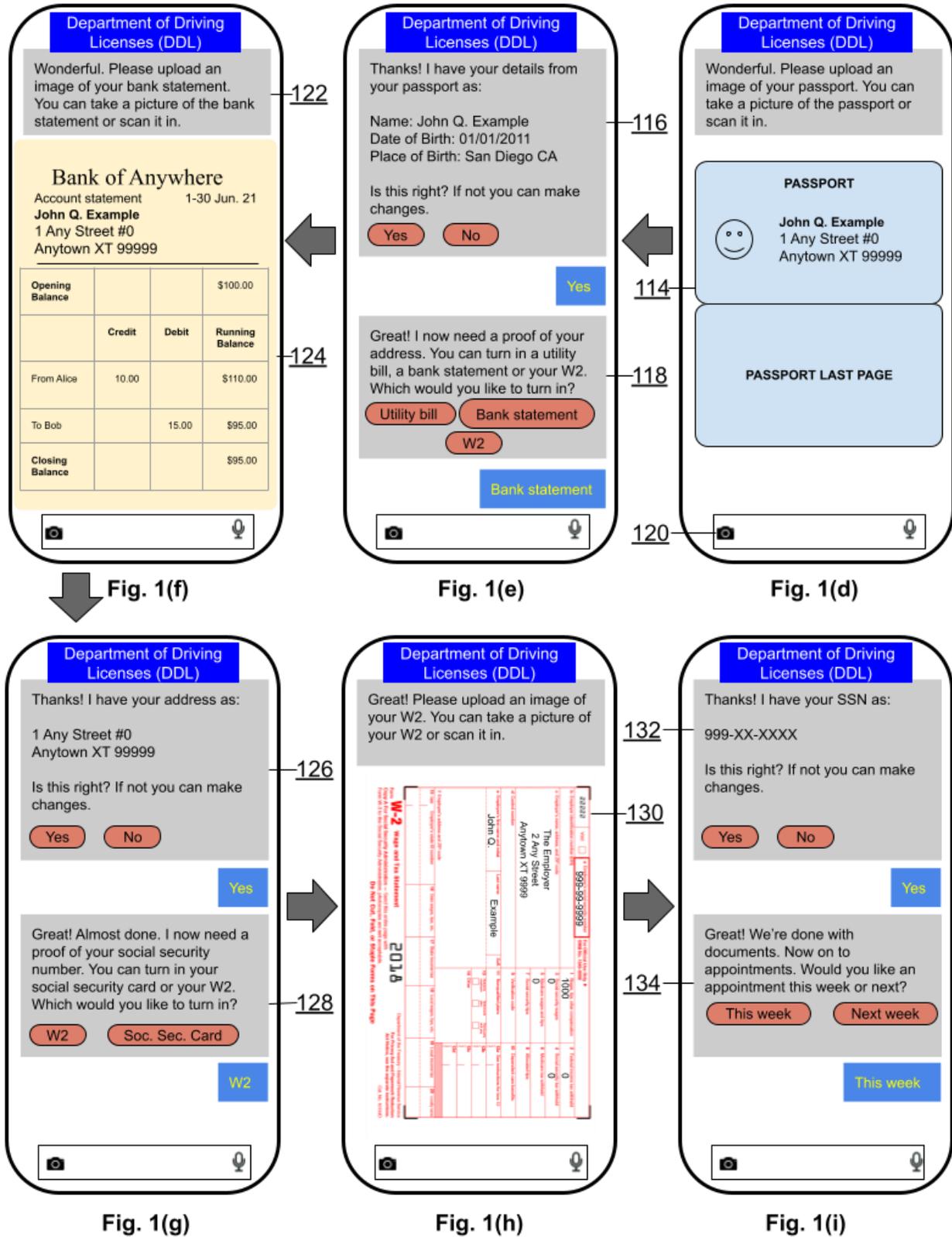


Fig. 1: Conversational document processing

Fig. 1 illustrates an example of conversational document processing, which applies conversational AI to process documents. A user opens an application provided by a licensing authority (Fig. 1a) to apply for and obtain a driver's license. A virtual agent (102) greets the user and, using natural conversation, offers to guide the user through the application procedure. The agent offers suggestions (104) and a quick-select set of likely answers to questions posed by the virtual agent.

The user selects the suggestion "schedule an ID appointment" (106). Alternative to selecting a suggestion, the user can also provide an answer using the microphone (110a) or type in the answer using a response field (110b). Based on the user's response, the virtual agent informs the user of the different documents needed for the particular application, e.g., to prove the user's identity and address (108). The user decides to use their passport as an identifying document. The virtual agent requests the user to upload an image of their passport (112).

The user uploads a scan of their passport (114), either by capturing a photo of the passport using a camera (120) or by uploading a scanned copy of the passport. The virtual agent parses the information in the passport and captures details found therein, e.g., name, date of birth, place of birth, etc. (116). The virtual agent presents the details to the user and requests the user to confirm their accuracy and manually provide any necessary corrections.

Based on the information thus far obtained and the information still needed to complete the driver's license application, the virtual agent guides the user in the process of submitting more documents (118), e.g., to verify the user's address. The user elects to use their bank statement to verify their address, and the virtual agent receives it (122). The user uploads a bank statement (124). The virtual agent parses the information in the bank statement and captures

details found therein, e.g., address (126). The virtual agent presents the captured details to the user and requests the user to confirm their accuracy and/or provide corrections.

Based on the information thus far obtained and the information still needed to complete the driver's license application, the virtual agent guides the user into submitting more documents (128), e.g., to verify social security number (SSN). The user elects to use their tax documents to verify their SSN and uploads a completed W2 form (130). The virtual agent parses the information in the W2 and captures details found therein, e.g., SSN (132). The virtual agent presents the details to the user and requests the user to confirm their accuracy and/or provide corrections. Having captured all necessary documentation and details therefrom, the virtual agent proceeds to make an appointment with the license issuing authority for the user (134).

The entire process of requesting and obtaining documents is performed securely and with user permission. The user can choose to provide specific documents or portions thereof. The user can elect to not have the virtual agent extract information and instead, enter it themselves. Data extraction from uploaded documents is done with specific user permission and in a secure manner. The documents are stored and redacted as appropriate, in accordance with the specific application and use context, user preferences, and applicable regulations. Here, application refers to the mobile, web, or other interface provided by an organization, e.g., a university, a licensing authority, etc., that accepts information from the user. Error handling is performed in real time. For example, if a document or fields within it are not successfully retrieved, the user is prompted to manually enter or correct data that is automatically read from the document. The timing of document submission and processing is automatically handled.

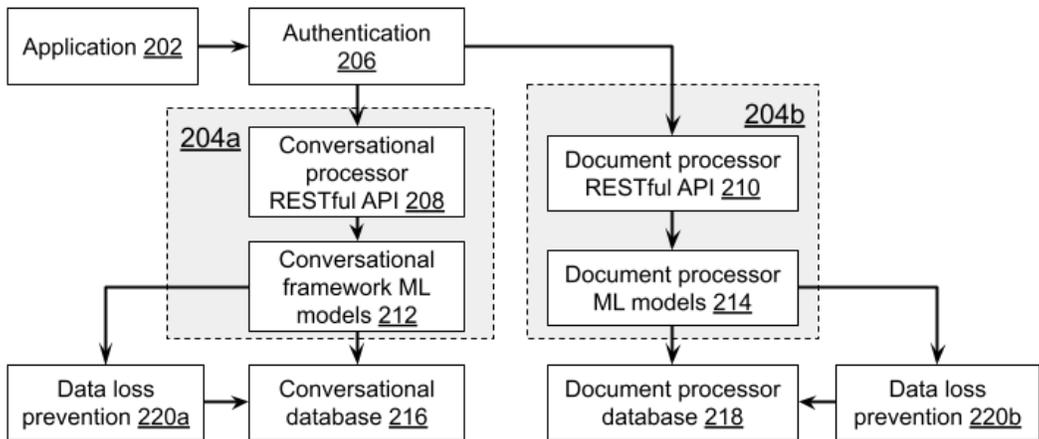


Fig. 2: Conversational document processing

Fig. 2 illustrates conversational document processing. Components to successfully guide users via a real time, humanlike, conversational interface to identify and submit digital copies of documents include:

- The *application* (202) is the (mobile, web, or other) interface of an organization e.g., a university, a tax authority, a licensing authority, etc. that accepts information from the user. The application invokes conversational document processing, e.g., the component subsets 204a and 204b, to obtain and verify information, including from documents uploaded by the user, in an efficient, frictionless, and conversational manner.
- The *authentication* (206) module authorizes access to conversational document processing. It can be implemented within the application.
- The *conversational processor RESTful API* (208) provides the application access to the conversational platform ML models for creating a humanlike conversation with the user that includes guiding (in a turn-by-turn manner) the user towards the appropriate documents to be submitted.

- The *document processor RESTful API* (210) provides the application access to the document processor ML models for parsing and extracting data from the submitted documents.
- The *conversational framework ML models* (212) are ML models used for natural language understanding, entity extraction, intent detection, and conversational turns.
- The *document processor ML models* (214) are ML models that parse and extract data from the documents submitted by the users. There can be different ML models for different types of documents. For example, there can be a bank processor model for bank statements; a tax-processor model for W2, 1099 forms; a driving license processor for driving licenses; etc.
- The optional *conversational database* (216) stores the conversation history between the user and the conversational framework ML models if permitted by the user.
- The optional *document processor database* (218) stores the data extracted by the document processor ML models if permitted by the user.
- The *data loss prevention* (220a-b) module provides users with the option of redacting sensitive information from the uploaded documents/images and the entities extracted from the documents.

In this manner, the described techniques of conversational document processing can, with user permission, obtain documents from a user and parse the content to extract information via a natural, humanlike conversational interface. A conversational interface as described herein can substantially expedite paperwork processes. Organizations operating in various domains such as financial services industry, education, tax and compliance, licensing, governments, etc. can provide conversational interfaces to their users to simplify document processing.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may

enable the collection of user information (e.g., information about a user's identity, documents, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that leverage conversational artificial AI to guide users through the process of identifying and submitting digital copies of documents for an application. A conversational interface as described herein enables organizations to use a single framework to address complex document processing use cases. The techniques can substantially simplify and expedite paperwork processes.

REFERENCES

- [1] <https://iona.ai/> accessed Nov. 26, 2021.
- [2] <https://iona.ai/onboard/media/images/video-20200918-131222-1b797529.mp4> accessed Nov. 26, 2021.
- [3] <https://quantiphi.com/> accessed Nov. 26, 2021.
- [4] <https://kore.ai/> accessed Nov. 26, 2021.
- [5] <https://cloudhub360.com/products/> last accessed Nov. 26, 2021.