Iasi City Explorer - *Alexa, what can we do today*?

Camelia Miluț
Faculty of Computer Science,
"Alexandru Ion Cuza"
University of Iasi
General Berthelot, No. 16
camelia.milut@info.uaic.ro

Adrian Iftene
Faculty of Computer Science,
"Alexandru Ion Cuza"
University of Iasi
General Berthelot, No. 16
adiftene@info.uaic.ro

Daniela Gîfu
Faculty of Computer Science,
"Alexandru Ion Cuza"
University of Iasi, Institute of
Computer Science, Romanian
Academy, Iasi Branch
General Berthelot, No. 16
daniela.gifu@info.uaic.ro

ABSTRACT

In everyday life, the voice communication always plays an essential role, being probably the most obvious feature that differentiates human being from other living entities. Due to its importance, speech recognition has become one of the most important applications of Artificial Intelligence (AI), highlighting how humans and machines interact. This paper describes a new voice application, developed as an Amazon Alexa skill, dedicated to the tourism area. More specifically, this system can contribute to improve the popularity of one of the most important cities from Romania, Iasi.

Author Keywords

Amazon Alexa; Amazon Echo; Speech recognition.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. H.3.2. Information Storage and Retrieval: Information Storage.

General Terms

Human Factors; Design.

INTRODUCTION

Over a few decades, speech recognition (SR) by machines is something that people have long dreamed of [9, 18]. Having a significant role in medicine, tourism, education, mobile computing, railway reservation, dictation, and web browsing, speech recognition systems (SRS) are still largely speaker/situation dependent [14]. SRS can be classified in several categories: according to speaker – *dependent*, *independent* or *adaptive* [13], according to situation (speech content) that includes various types of – *speech utterance*, *vocabulary*, *channel* [17]. Even if the progress is significant, the scaling any speech recognition system has always been a significant obstacle, especially in the travel industry.

The main research question of this paper intends to answer: *How efficient is an automatic voice system in the tourism area to offer practical uses to different peoples categories?*

Our survey proposes a new voice application, developed as an Amazon Alexa skill, dedicated to the tourism area (here, Iasi). Precisely, this system, Iasi City Explorer, can be used to give users (travelers, people with vision problems, etc.) information regarding their trip. Voice searching can suggest recommendations for restaurants, cafes, car rentals, information about weather and activities suitable for a certain time of the day, including main attractions and places where you can pleasantly spend your time.

The paper is structured as follow: Section 2 presents a short overview of voice assistants, starting with Alexa, in order to clarify their importance and what can we do to add new functionalities to improve our daily lives, while Section 3 refers to the architecture and the design structure of Iasi City Explorer system, Section 4 briefly discusses the evaluation of this application, comparing two sets of participants according to their age before drawing some conclusions in the last section.

GUIDE TO VOICE ASSISTANTS

The history of voice assistants and devices used in human–computer interaction (HCI) is not recent. In 1961, IBM introduced Shoebox[1], first SRS. It could recognize 16 words spoken by the built-in microphone and digits from 0 to 9. Also, it performs simple math operations. In 1962, its functionalities were demonstrated by William C. Dersch, the developer.

At today, a voice assistant, known as the *smart speaker*, is a software product that is activated by voice interaction. This software can assist people with different tasks. Besides Amazon, other companies such as Google, Apple, etc. have launched such products. Below, these systems will be presented briefly.

Alexa

It is a voice assistant, being, also, a platform like Android or IOS[2]. Alexa's name appeared for two reasons. First, because of the consonant X found in name, the sound of

[1] https://onlim.com/en/a-short-history-of-voice-assistants-with-alexa/

[2] https://onlim.com/en/a-short-history-of-voice-assistants-with-alexa/

which, the name does not resemble those of similar devices, to prevent their accidental activation. Second, for the connection to the Library of Alexandria - this was considered to be the keeper of the knowledge of all time [Fagna, 2017].

Like other systems, Alexa has applications (called skills) that can extend its use. They come in a variety of categories, including areas such as financial, news, productivity, weather, music, etc.

Basically, these skills make Alexa *smart* [3]. These are functionalities that Alexa can send us reminders, set alarms, can program smart devices in the house so that the light is optimal, start or stop at a certain time. It can also do things in our place, such as giving pizza order, starting music, posting on Twitter, etc. Alexa's skills are on the rise, reaching today over 30,000 skills available to US users in March 2018. [8].

It is integrated with many physical devices such as smart home appliances (refrigerators, bulbs, thermostats, etc.), automotive, surveillance cameras, locks, mobile phones, but the most used are the Echo smart speakers launched by Amazon in 2014 [7].

Alexa listens to the voice commands received from the user and provides an appropriate response through which it performs a task [6]. Alexa for mobile devices allows the user to activate the embedded device by a keyword or by clicking a button [3, 5].

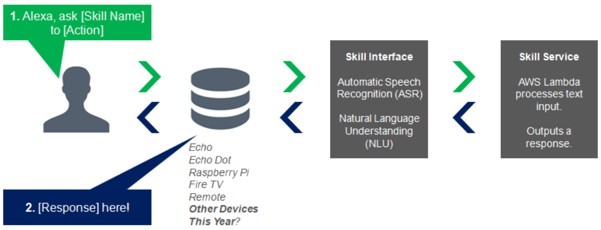
1. Alexa, ask [Skill Name] to [Action]
2. [Response] here!
Echo
Echo Dot
Raspberry Pi
Fire TV
Remote
Other Devices
This Year?
Skill Interface
Automatic Speech Recognition (ASR)
Natural Language Understanding (NLU)
Skill Service
AWS Lambda processes text input.
Outputs a response.

Figure 1. Operating scheme of an Alexa skill [3].

Switching from 25,000 skills to Alexa skill store to 30,000 it lasted 97 days. This means an average of 51.5 new published skills per day!

Amazon Echo

It was the first device built with Alexa launched on the market [7]. Initially, it was presented as an intelligent box through which you could control the music you are listening to, through voice interactions, and a few other things. Over time, Alexa's skills have evolved considerably. Today, the most popular devices controlled through Alexa are found in the smart home: Philips Hue's light bulbs, Cloud Cam - an indoor security camera system, Ecobee 4 - smart thermostat and so on.

Google Assistant

We find it on most Android devices as an operating system. It was originally thought of as an extension of "*OK Google*" - the existing voice control service. Google Home is the smart speaker associated with this digital assistant. This device doubles as a command center for a smart home and personal assistant for the entire family to help with day tasks with more ease[4] [16].

Siri

Apple proposed Siri as a digital assistant, being released with the iPhone 4S in October 2011. This assistant ran from a project of the SRI International Artificial Intelligence Center. Its voice recognition engine was provided by Nuance Communications, a vocal recognition company. Siri uses advanced learning technology. Voice recordings for this assistant began in 2005, with actors not knowing exactly the purpose of those recordings.

Currently, Siri is present on all devices that use iOS, watchOS, macOS, or tvOS as operating systems. Apple HomePod is Apple's smart speaker version. It brings together Apple Music and Siri to learn the musical tastes of the user. It can also control smart home devices and can do daily task management.

Cortana

In 2009, Microsoft began the development of Cortana, a voice assistant system [10]. Its launch took place in April 2013 at the Microsoft BUILD Developer Conference. The Cortana is integrated on the Windows operating system, both on the PC and on the phones. Android and iOS apps are also available. The Cortana works with both voice and text interaction. This digital assistant can set alarms, make appointments, and search on Bing. It is connected with third party applications such as Netflix or Audible. Behind Cortana there are automatic learning algorithms to provide a personalized experience for the user. Microsoft has not yet launched a smart speakerphone for its voice assistant. However, Cortana is not absent in the smart home hubs. Currently, the digital assistant is supported by a German company called Harman Kardon, which produces intelligent speakers integrated with it.

Similar skills

Skills for Alexa, similar to Iasi City Guide, are:

- *Bucharest Guide*[5]: offers information about 33 points of interest in Bucharest.
- *Rome tour*[6]: A guide to tourist attractions in the capital of Italy. It contains the most popular locations in Rome,

[3] https://www.zldoty.com/tag/alexa/

[4] https://www.youtube.com/watch?v=XVjyIA3f_Ic

[5] https://www.amazon.com/Catalin-Batrinu-Bucharest-Guide/dp/B074WBNM7R

such as monuments or museums. Skill sends location addresses to companion application on mobile or TV.

- *NYC Guide*[7]: Provides information to tourists in New York such as tourist attractions or information about transportation.

- *MyParis Guide*[8]: provides detailed information about the top 5 things to do in Paris.

PROPOSED SOLUTION

The main purpose of *Iasi City Explorer* is to enhance the tourist experience in Iasi and also to help the newcomers to explore the city and find location easier, without the need to make intermediate search.

Iasi City Explorer is available on the Amazon Echo device, but also on web browsers, Android and iOS platform, via Amazon Alexa app, where the user can interact with the application through voice commands. The main reason for picking Alexa as a development platform is the increasing popularity of the Artificial Intelligence, implicitly of the digital assistants and *voice first* applications.

Architecture

At the development level, the application follows a cloud-based architecture [1], orchestrating services provided by Amazon Web Services[9] with Google Maps Platform[10] and Yahoo Weather[11]. The interaction between services can be seen in the Figure 2.

The user interacts with Alexa-integrated devices such as Amazon Echo, mobile applications, etc. The query is intercepted by Alexa Voice Services, and is forwarded to the Alexa Skills Kit that launches the Lambda function [12, 15]. Depending on the request received, Lambda takes data from Yahoo Weather API, Google Places API or DynamoDB[12]. DynamoDB data is also obtained through the Google Places API. To monitor the Lambda function, CloudWatch metrics are recorded.

To access DynamoDB and CloudWatch, the Lambda function gets some permission through Identity and Access Management (IAM). The response processed using the Lambda function is passed back to the user through the Alexa Voice Service and the device through which the interrogation was made.

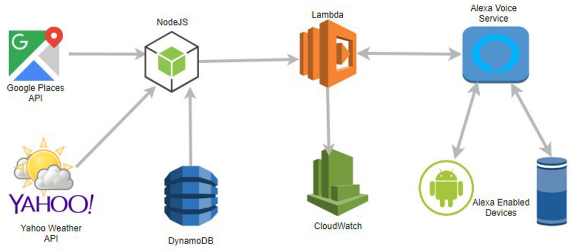

Figure 2. System architecture.

In the following, we will present the technical details of this process.

Developing the Skill

An Alexa skill is composed of two parts:

- *Skill interface*: The user interaction side, configured by Alexa Skill Console. This section defines how the user's voice commands are directed to the Skill Service.

- *Skill service*: contains the logic of the application, hosted on a remote server.

Skill Interface

Skill Interface is the party responsible for processing the spoken language. It deals with the translation of voice input from the user in events that can be processed by Skill Service. The interface sets the name by which the skill can be invoked (also called invocation name). For example: "*Alexa, open Iasi City Explorer*". "*Iasi City Explorer*" is the name through which the skill is identified.

To know how to listen to words spoken by the user, Skill Interface uses an interaction model. The developer defines the words that are mapped by certain intents in the interaction model, giving a list of words (utterances). This list of words associated with intent suggests the different ways the user can interact with that skill. These are used to generate the natural language processing model.

Intent Scheme

Intent is an indicator by which the user launches a particular command. The Intent Scheme is a JSON object that contains the names of all the intentions the application can handle, as well as the word lists that trigger that intention and the list of slots, where appropriate.

Alexa Skill Kit offers a number of predefined intentions. Also, the developer can define a series of customized intentions. The customized goals defined in City Explorer are as follows:

- *AboutIntent*: to get a brief description of the city. Launched by the expressions "*about*" and "*tell me about this place*".

[6] https://www.amazon.com/Chris-Cinelli-Rome-tour/dp/B07475HGSS

[7] https://www.amazon.com/AlexSantisteban-New-York-City-Guide/dp/B076QGS2PS

[8] https://www.amazon.com/Adrien-Chan-MyParis-Guide/dp/B01MT2T4O6

[9] https://docs.aws.amazon.com/IAM

[10] https://developers.google.com/places

[11] https://developer.yahoo.com/weather/

[12] https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Programming.html

- *AttractionIntent*: to receive a tourist attraction recommendation and learn about it. Triggered by the words: "*recommend an attraction*", "*give me an activity*", "*what to visit*", "*tell me about a place to visit*".
- *FoodIntent*: Restaurant recommendations based on a preference or generic recommendation, if preference is not specified. Triggered by expressions containing the "*dish*" slot, for a specific recommendation: "*dish*", "*where can I get some dish*", "*I would like some dish*", and for a general recommendation "*food*", "*I want to eat*".
- *ActivityIntent*: Suggests the user a place to perform a certain activity (e.g. *I want to swim*). Examples of phrases: "*recommend location*", "*suggest location*", "*I want location*", "*I want location*", "*give me a location*".
- *CarIntent*: Suggest companies that can rent cars. Activated by the expressions "*rent a car*", "*I would like a car*".
- *RecommendIntent*: location recommendations based on time of day. Triggered by: "*what should I do*", "*where should I go*", "*give me an idea*", "*give me something to do*".
- *GoOutIntent*: Time and weather. Turned out by: "*go out*", "*go out-side*", "*how is the weather*", "*weather in Iasi*", "*weather*", "*what time is it*".

Utterances

An utterance is a word or expression attributed to an intention to trigger it. The list of utterances helps the Alexa Skill Interface process the words spoken by the users in intentions.

Slots

A slot represents a variable name, which is assigned by the user during the speech. Slots used in the application are built-in, which means they are offered by the Alexa Skills Kit. City Explorer uses 2 slots, namely `AMAZON.Foods`, and location - like `AMAZON.LocalBusinessType`.

These two types of slots include dishes (e.g. *chocolate, cake, scrambled egg, Campbell's low sodium chicken broth*) and a list of location types (*medical clinics, food store, auto rental store, dry cleaning*).

Skill Service

Skill Service implements event handlers that define how skill will behave when the user triggers an event through a specific skill question.

Configuring the Lambda function

The role of Lambda's skill is to interpret user interaction and provide an adequate response. This function runs every time the user interacts with the application. The function receives as input a JSON containing the processing of the voice command received from the user, made through the interaction model.

The programming language used to develop the Lambda function was JavaScript, with Node JS. This language has been chosen because support for AWS SDK is provided and is recommended for the development of small server-side applications where asynchronous operations are required.

To create a Lambda function, it is necessary for the developer to have an AWS account. The next step is accessing the console on aws.amazon.com and selecting Lambda from Compute Services. To work with the Alexa Skill Kit, the Lambda function must have the endpoint in US East (N. Virginia) or EU (Ireland). Next, the developer chooses a running environment for function `LocalRecommendation` (the name of Lambda function used in *Iasi City Explorer*), this is Node.js, 6.10 and a blueprint to have everything set for alexa-sdk. For the Alexa Skills Kit to trigger the Lambda function, it must be set as a trigger.

Deploy

To provide the necessary permissions, the developer must create an IAM role. This role includes permissions to write in AWS CloudWatch and read information from DynamoDB. After setting up the role, it is attached to the function. For the skill created to be functional, Lambda must be available for Alexa Voice Service. This can be done directly, if the code is written in the editor provided by AWS Console. If the code is deployed in a local environment, it can be loaded into an AWS Simple Storage Service Bucket (AWS S3) and provided a link to it, uploaded as an archive in AWS consoles or from the command line through AWS Cli. These can be made available via the "*ask deploy*" command.

Handling Intents

Manipulation of intents and processing of Alexa's answers is done through the Lambda function. The *Iasi City Explorer* application contains seven intents, through which the user can learn more about the city of Iasi and the activities it has to offer. In addition to customized intent, the application also uses predefined intentions from the Alexa Skills Kit (*HelpIntent*, *YesIntent*, *NoIntent*). Next, we will describe how the features described in the intent schema were obtained.

User-initiated questions to the application are translated by Voice Service into intents. These are two-way *with slots* and *without slots*.

For intents *with slots*, Lambda makes a request to Google Places to generate the response, depending on the content of the slot. The desired location type is transmitted as a search text, and from the list provided as a result, a location in the top five is randomly selected. It then filters the location information and provides the voice and visual output (in the companion application) accordingly.

For intents *without slots*, the necessary data is stored in a database. Depending on the type of location required for the intent, a query is made to the DynamoDB database. The result of the query is properly processed and transmitted as voice and visual output.

Extracting Data

To populate the database, a separate section of the application was created. This section includes a Node JS server that contains a DynamoDB client linked to the *PLACE LIST* table and a list of location types made by Google Places queries through Place Search. In order to get more information on the locations obtained from the query, they are provided as a Place Details function. The results are filtered and serialized in JSON format. Finally, serialized results are inserted into the table through the DynamoDB client.

USABILITY TESTING

In order to provide both a real analysis and a range of applications that require users' feedback, a series of usability tests was performed. The targeted users were belonged to two age categories: the elders (here, with visual deficiency) and youngsters. The latter category participants were blindfolded to better impersonate people that have no mean of using visual aid [2].

Methodology

The performed test was composed of a presentation regarding the problem context, several tasks to be conducted during the experiment, a satisfaction questionnaire and a brainstorming session to gather ideas for other useful functionalities. Each participant interacted with the assistant in a different environment, without communicating with other users prior to the experiment. These measures were taken with the purpose of obtaining mostly unbiased results. Everyone had to perform on the provided smartphone application 5-6 tasks, i.e. voice interactions with the application (from the set "*tell me about this place*", "*recommend an attraction*", "*Where can I get some dish*", "*recommend a location*", "*I want to rent a car*", "*give me something to do*" and "*weather in Iasi*"), which require approximately 3-4 minutes per session.

Once they finished the tasks, the overall experience was assessed by using SUS (System Usability Scale) [4], which presented us with the results described in the following sections. The brainstorming session was organized in groups of three to five people from the same age sector, in order to better understand and assess which functionalities would better fulfill their general needs.

Elders with visual deficiency

Participants: We collaborated with a group composed of 4 people who have little experience with technology. Their age was between 50 and 56 and all of them experienced various eyesight deficiencies. Regarding the usage of smartphones, 100% percent of the participants have used a smartphone before, while only 50% interacted with a smart assistant (e.g. Siri, Google Assistant).

Results: As a first impression, the participants enjoyed the simplicity of interaction with the application, mainly because the authentication method is fast and reliable, there are no intricate interface elements like small buttons or text input fields and the vocal interaction substitutes entirely the need for glasses or fairly good eyesight.

After the proposed test scenario was executed successfully, each of them was asked to rate their general experience with a grade from 1 to 10, where 1 stands for confusing/ frustrating experience, and 10 for clear/pleasant experience. The average rating for the user experience was 7,75, since three of our participants were not fluent English speakers, which often resulted in misunderstanding of the responses to some part of the vocal enquiries. This questionnaire was followed by a feedback and brainstorming session meant to highlight the current issues and the most desired future improvements. We concluded that an increase in the collection of available languages was a development priority in the near future, alongside with the possibility to repeat answers in case of need.

Young people without visual disabilities

Participants: In order to assess the opinion of another age segment, we repeated the experiment with 6 student peers, aged between 18 and 23, which can be categorized as experimented technology users. All of them use their smartphones on a daily basis and require the help of Iasi Smart City application for easy tasks at hand. Since we needed to simulate that the proposed use case is performed by people with visual deficiencies, this segment of participants used blindfolds or similar methods to take full advantage of the applications features.

Results: Analyzing the gathered preliminary observations, the participants did not seem to experience difficulties while performing the requested tasks. Even though the smartphone usage experience was opposed to the usual interaction, they found enjoyable that the level of access to information remained similar to the one they were accustomed to.

After completing the test scenario, they were asked to rate their general experience with a grade from 1 to 10, similar to the previous set of participants. The average rating for the user experience was 8, since young people are more inclined to embrace technology. When asked for their feedback and desired future features, most of them opted for adding options for social media access, as well as the possibility to send and read text messages. These improvements would surely increase the popularity of the application among this segment of the population.

Remarks

When comparing the two sets of participants, we noticed a certain reticence among the elder group in regard to the usage of technology in completing their daily tasks. Nonetheless, the youngsters seemed to be far more receptive to the idea of allowing a smart application to help them in any situation they might find themselves into.

To make the evaluation process more transparent, we chose to use SUS [4], in which the responses were multiplied by 2.5, thus obtaining a scale from 0 to 100 from the original 0 to 40 scores. These are considered to be percentile ranks [11].

Therefore, we concluded that Iasi City Explorer application can support and help people with visual deficiencies in fulfilling their activities when they visit Iasi city. Also, the application would have a great impact among all ages, regardless the severity of their condition.

CONCLUSION

Alexa was chosen as the application development platform because *voice first* applications are becoming increasingly popular, and one of the reasons is ease of use. An application that can be used by voice interaction can be easily used by visually impaired people or with motor disabilities or by people who are reluctant to interact with a touchscreen (the elderly, or children). Also, Alexa as a development platform is portability. Besides the popular Echo devices, Alexa is integrated with many IoT devices, but also with classic smartphones, making it accessible at all times and in any place. Moreover, Iasi City Explorer system can contribute to improve the popularity of one of the most important cities from Romania. We can adapt the system to users' information needs.

ACKNOWLEDGMENTS

We thank all the volunteers involved in the development of this application. This work is partially supported by POC-A1-A1.2.3-G-2015 program, as part of the PrivateSky project (P_40_371/13/01.09.2016).

REFERENCES

1. Alboaie, L. Servicii/Furnizori de servicii in Cloud. (2019) https://profs.info.uaic.ro/~adria/teach/courses/CloudComputing/resources/CC4_CloudComputing.pdf
2. Calancea, C. G., Miluț, C. M., Alboaie, L., Iftene, A. iAssistMe - Adaptable assistant for persons with eye disabilities. In *23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2019)*, 4-6 September, Budapest, Hungary (2019)
3. Claude, J. P. Developing Alexa Skills. *The Big Nerd Ranch Guide* (2016)
4. Drew, M., Falcone, B. and Baccus, W. What does the System Usability Scale (SUS) Measure? *Springer International Publishing*. (2018)
5. Fagna, A. Understanding Amazon's Alexa and Building Alexa Skill. *medium.com*. (2017)
6. Filimon, M., Iftene, A., Trandabăț, D. Bob - A General Culture Game with Voice Interaction. In *23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems* (*KES 2019)*, 4-6 September, Budapest, Hungary. (2019)
7. Johnson, B. How Amazon Echo Works. *HowStuffWorks*. Tech. Electronics. Gadgets. High-Tech Gadgets (2016)
8. Kinsella, B. Amazon Alexa Skill Count Surpasses 30,000 in the U.S. *voicebot.ai*. (2018) https://www.voicebot.ai/2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/
9. Milut, C., Iftene, A. and Gifu, D. Iasi City Explorer. In: Proceedings of the 5th Conference on Mathematical Foundations of Informatics, MFOI-2019, Gifu, D., Aman, B., Iftene, A, Trandabat, D. (eds.), pp. 271-275 (2019)
10. Ravenscraft, E. Everything You Can Ask Cortana to Do in Windows 10 (2015) https://lifehacker.com/everything-you-can-ask-cortana-to-do-in-windows-10-1721725525
11. Rogosa, D. Accuracy of Individual Scores Expressed in Percentile Ranks: Classical Test Theory Calculations. *Stanford University* (1999)
12. Rouse, M. AWS Lambda (Amazon Web Services Lambda). *Searchaws.techtarget* (2019) https://searchaws.techtarget.com/definition/AWS-Lambda-Amazon-Web-Services-Lambda
13. Saksamudre, S.K., Shrishrimal, P.P., Deshmukh, R.R. A Review on Different Approaches for Speech Recognition System. In: International Journal of Computer Applications, Vol. 115, No. 22: 23-28. (2015)
14. Sherman, W.R., Craig, A.B. Interface to the Virtual World-Input. In: Understanding Virtual Reality (2003).
15. Tejada, Z., Wilson, M., Buck, A. and Wasson, M. Non-relational data and NoSQL. *Microsoft Azure*. (2018) https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data
16. Tillman, M. and Gragham, D. What is Google Assistant and what can it do? *www.pocket-lint.com* (2019)
17. Vimala, C. and Radha, V. A Review on Speech Recognition Challenges and Approaches. In: World of Computer Science and Information Technology Journal (WCSIT), Vol. 2, No. 1: 1-7 (2012)
18. Zhang, Z. Mechanics of human voice production and control. In: J Acoust Soc Am. 140(4): 2614-2635 (2016).