FASSI 2017

The Third International Conference on Fundamentals and Advances in Software Systems Integration

ISBN: 978-1-61208-587-6

September 10 - 14, 2017

Rome, Italy

FASSI 2017 Editors

Chris Ireland, Open University, UK

Luigi Lavazza, Dipartimento di Scienze Teoriche e Applicate - Università degli Studi dell’Insubria, Italy

FASSI 2017

Forward

The Third International Conference on Fundamentals and Advances in Software Systems Integration (FASSI 2017), held between September 10-14, 2017 in Rome, continued a series of events started in 2015 and covering research in the field of software system integration.

On the surface, the question of how to integrate two software systems appears to be a technical concern, one that involves addressing issues, such as how to exchange data (Hohpe 2012), and which software systems are responsible for which part of a business process. Furthermore, because we can build interfaces between software systems we might therefore believe that the problems of software integration have been solved. But those responsible for the design of a software system face a number of trade-offs. For example the decoupling of software components is one way to reduce assumptions, such as those about where code is executed and when it is executed (Hohpe 2012). However, decoupling introduces other problems because it leads to an increase in the number of connections and introduces issues of availability, responsiveness and synchronicity of changes (Hohpe 2012).

The objective of this conference is to work toward on understanding of these issues, the trade-offs and the problems of software integration and to explore strategies for dealing with them. We are interested to receive paper from researchers working in the field of software system integration.

We take here the opportunity to warmly thank all the members of the FASSI 2017 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to FASSI 2017. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the FASSI 2017 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that FASSI 2017 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of software systems integration. We also hope that Rome, Italy provided a pleasant environment during the conference and everyone found some time to enjoy the historic charm of the city.

FASSI 2017 Chairs

FASSI Steering Committee

Chris Ireland, The Open University, UK
Hironori Washizaki, Waseda University / National Institute of Informatics / System Information, Japan
Keijiro Araki, Kyushu University, Japan

FASSI 2017
Committee

FASSI Steering Committee

Chris Ireland, The Open University, UK
Hironori Washizaki, Waseda University / National Institute of Informatics / System Information, Japan
Keijiro Araki, Kyushu University, Japan

FASSI 2017 Technical Program Committee

Harvey Alférez, Montemorelos University, Mexico
Keijiro Araki, Kyushu University, Japan
Graeme Burnett, University of Glasgow / Temenos Ltd, Scotland
Leire Etxeberria Elorza, Mondragon Unibertsitatea, Spain
Ip-Shing Fan, Cranfield University, UK
Alan Hayes, University of Bath, UK
LiGuo Huang, Southern Methodist University, USA
Chris Ireland, The Open University, UK
Slinger Jansen, Utrecht University, Netherlands
Tomi Männistö, University of Helsinki, Finland
Anh Nguyen-Duc, Norwegian University of Science and Technology (NTNU), Norway
Tosin Daniel Oyetoyan, SINTEF Digital, Norway
Christian Percebois, University of Toulouse, France
Olivier H. Roux, Ecole Centrale de Nantes, France
Gunter Saake, Otto-von-Guericke-University of Magdeburg, Germany
Hironori Washizaki, Waseda University / National Institute of Informatics / SYSTEM INFORMATION, Japan
Tim Weilkiens, oose Innovative Informatik eG, Germany

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Hashtag Trend Analysis based Travel Destination Recommendation Technique

Hyunsoo Hong, Sungwook Yoon, Hyenki Kim
Dept. of Multimedia Engineering
Andong National University,
Andong-City, Republic of Korea
e-mail: bazooka92@nate.com, uvgotmail@nate.com, hkkim@anu.ac.kr (Corresponding Author)

***Abstract*—The study explores a plan for the evaluation of travel destination preferences classified by emotional language by analyzing hashtags existing in image sets through the collection of Social Network Service, and by comparing with the vocabulary showing general travel satisfaction. The vocabulary to be evaluated was divided into regional, tourism target vocabulary, common expression language, and emotional expression language. This study demonstrates tourists' preference for travel destinations according to whether collected hashtags match previously retrieved emotional hashtags when tourists search for specific tourist sites. Through this, it will be possible to analyze the changing travel trends of a specific travel destination and accordingly, it will be possible to establish a travel promotion policy.**

Keywords-hash-tag; trend; emotional; corpus anaysis.

I. INTRODUCTION

Through tags, most of the social network service (SNS) platforms tend to emphasize key words that express their writing and central key words. The role of keywords, categories, and metadata is expanded to express self-directed information discovery on digital documents, which is expressed as an additional expression with a hashtag. This resulted in the users of the existing metadata transferring from experts to general users. Through the hashtag processing, the folksonomy-based voluntary classification of ontologies allows various processes for corpus discovery of big data to be arbitrarily organized and used as search and analysis data [1]. The frequent use of hashtags in social network services is common to many platforms, and it is possible to intuitively perform statistical analysis, such as association analysis or frequency analysis between hashtags. Using this information, the promoter can use the hashtag to actively promote tourism information [2].

In general, hashtags are usually expressed by summarizing articles written by SNS users into noun-type keywords, and are responsible for categorizing keywords searched for information sharing.

The association analysis and the real-time pattern analysis of the promotion keywords are considered to be the means to make or strengthen tourism trends. This method of promoting tourist attractions is also used as a means of promoting tourism information to tourists. Municipal tourism information from municipalities with tourist attractions generally promote events, festivals, promotion of specific areas, Fam tour, and local companies, and it is intended to spread through the use of not only official channels but also platforms, such as Facebook, Instagram, Twitter, Kacao Story, Band, and Blogs. Typically, these platforms are made through social connections with users and organizations, so they spread even faster through related tags.

The acquisition of information on tourist sites tends to be affected by the names of the sites and the keywords of the empirical contents posted on social platforms. The emotional and abstract experience language in the travel review, along with the name of the locality posted on social media will possibly affect the next trip destination of followers and friends exposed to the media [3].

However, since social media generally involves a rapid change in trends, it reflects the trend of recent time changes in real time. In general, the analysis of these trends reflects the tendency to change through folksonomically categorized tags. This reflected travel trend influences search preferences and strengthens or damages the emotional image along with the place's name. This research aims to provide effective promotional strategies for tourist information on targeted social network service tourist sites. Some methods are proposed, which would help to collect retrieved hashtags and select destinations in accordance with travel trends changing in real time. The study conducted research on the effective use of multiple hashtags as keywords for travel through classification and analysis.

The structure of this paper is as follows. Session 2 describes related studies, Session 3 describes tourism travel destination preference analysis plan and Session 4 describes the conclusion of the paper.

II. RELATED STUDIES

Articles posted on social networks with the subject of travel are always being uploaded from the point of view of tourism consumers, and the followers of digital connections recognize this information to be trends. Users who are interested in traveling can learn information about the desired travel destination through hashtag searches and can enhance the travel trend. A technique [4] was also proposed where, the more the travel destination name is gathered on Twitter, the more it is considered to be the travel destinations that people are likely to visit. In addition, there was a study [5] that analyzed the real-time SNS trends to recommend travel destinations and a study [6] on techniques of analysis of hashtags that appear simultaneously in the SNS environment. This study suggests the level of

Copyright (c) IARIA, 2017. ISBN: 978-1-61208-587-6

recommendation according to the survey done on emotional vocabulary and the frequency analysis of existing hashtags. Unlike previous research, it is to grasp the preference through the emotional analysis through collective intelligence.

A. *Folksonomy based Hashtag*

Folksonomy is a new classification system based on 'tags' instead of 'directory' which is the traditional classification standard, and it means 'classification by people' (Folk + order + nomos). According to Wikipedia, "it is a newly coined word that represents the way members organize information together using freely chosen keywords" [7]. In addition to dividing into politics, economy and culture, for example, it is possible to gather information at a glance based on keywords, such as general elections and flower festivals. The difference between folksonomy and taxonomy, which is a classification system like existing book management, is that the members voluntarily organize unit information by giving meaning to individual information, and since the information is listed while interacting with the main members, it has the advantage of transmitting more meaningful information more accurately, and therefore it is in the same context as the hashtags on SNS [8]. Therefore, folksonomy is not as systematic compared to taxonomy. This is because users of the Internet are not making and learning a folksonomy system, but making it more convenient and useful. Thus, the cost of categorizing content can be reduced because there is no need to learn a complete classification scheme. Folksonomy is expandable and can change quickly depending on how users are categorizing through the Internet. Since folksonomy is the result of a collection of people's interaction, it is suitable for searches. The folksonomy-based hashtag also needs to be evaluated for changes in its trends and categories.

B. *Status of Use of Promotion tags to Promote Municipal Government*

In April 2017, the study conducted a questionnaire survey on the selection of promotional phrases related to Yeong-ju city to about 1,000 public officials who were all the subjects of public promotion training for Yeong-ju city.

Figure 1 shows it processed by word cloud according to frequency.

Figure 1. Word cloud based on frequency of promotional keywords

In the top 40 keywords for promotion, 90% of language bundles, 5% of common language, and 5% of emotional corpus showed uneven usage combinations. The name of the place or specific sightseeing spots were mainly surveyed. Figure 2 below shows the results of the frequency analysis.

The results show that according to the classified keywords used by tourism promoters, there is a tendency to promote geographical names. (see Figure 2).

BuseokTemple	105	Sosuseowon_Confucian_Academy	104	apples	72	Sobaeksan_Mountain	65
the	64	ginseng	56	museom	33	Seonbichon	31
Punggi	28	Seonbi	25	museo	24	village	23
Insam	21	Floating_Stone	20	rayon	20	Korean_beef	16
town	16	apple	15	beef	15	BuseokTempl	15
of	15	decimal	13	in	13	Yeongju	11
Healing	10	si	10	bridge	9	forest	9
grapes	9	healing	9	log	9	Minority	9
Yangban	9	Confucian	8	Ginseng	8	Museom	8

Figure 2. Promotion keywords based on survey

The hashtags used on SNS are typically used as multiple lists, and it is common to be listed mainly through the placement of the area name - the general corpus - the common corpus. This study aims to gauge the level of travel satisfaction by combining names of certain areas, common expression language containing the contents of tourist experiences, and corpora of folksonomical emotional expressions.

III. TOURISM TRAVEL DESTINATION PREFERENCE ANALYSIS PLAN

A. *Hashtag Classification Technique*

This study proposes an analysis method that calculates a tourist travel destination preference index by using SNS data created by tourists on travel destinations and analyzing the frequency of specific tourism travel destination references and emotional analysis. The contents of the hashtag used as a keyword for promotion were classified into three categories in this study.

A list of words in the order of geographic, common, and conceptual words is shown in Figure 3.

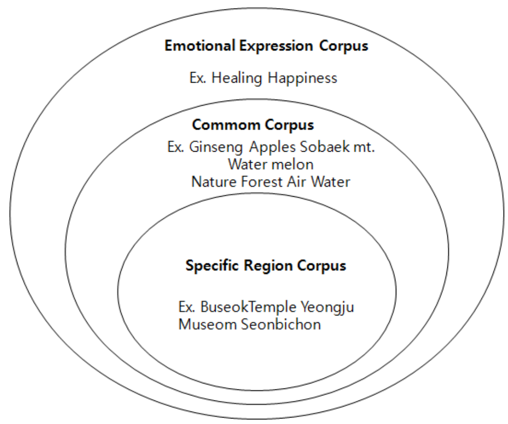

Figure 3. Travel purpose and travel destination classification

They were divided into {specific region corpus, common expression corpus, emotional expression corpus}. In this classification, the study created a database with a real-time hashtag trend list of emotion expression corpus (Table 1).

TABLE I. CORPUS-SPECIFIC PERSONALITY REGULATION

	Explanation	**Example**
Specific Region Corpus	Name of tourist destination, geographical name, tourist destination	BusekTemple Jeju
Commom Expression Corpus	Non-border language, such as natural environment	Mountain, Sea, Apples
Emotional Expression Corpus	Abstract word representing travel purpose	Healing Happiness Memories

B. *Analysis Plan*

The flow chart of the proposed method is shown in Figure 4.

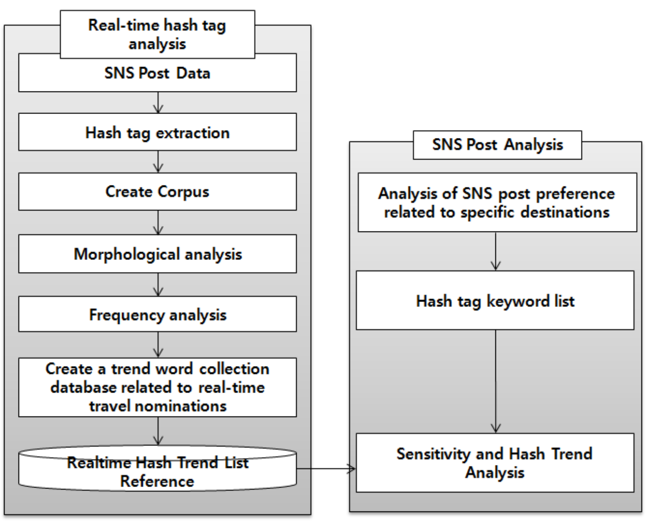

Figure 4. Flowchart of Research

First, collect hashtags from posted writings on SNS. Then, analyze morphemes through corpus analysis. Finally, collect related hashtags used in the same sentence. Analyze articles containing representative keywords related to various travel location names through hashtag collection of posts on social network services. The study uses Korean morpheme analyzer [9] to classify articles of social platforms, such as Instagram, Facebook, and Twitter into noun units and extract candidate keywords. Then, the term frequencies in which the keywords appear are listed in order. In this case, words not relevant to the research purpose are filtered by Stop Word. In addition, to identify the real-time trend area, words related to the region and emotional words are extracted. Finally, compile only the keyword related to the area and list the ranking. The items with highest frequency have priority.

A name dictionary for tourism travel destination is constructed in advance, and based on this, SNS including a tourist travel destination name is extracted, and through the term frequency algorithm, the frequency of related keyword references within the SNS text of each travel destination is obtained. In addition, emotional words for sightseeing travel destinations are extracted as keywords, and a trend dictionary database of emotional vocabulary collected from posts about travel is created.

$$P(t) = F_t\left(\sum_{n=0}^{\infty} HT_n + V_n\right) \qquad (1)$$

Equation (1) is the preference index (*P(t)*). *t* is a tourist destination, *Ft* is the total number of *t* references, *Ft* is the reference frequency of *t*, *n* is the number of posts including *t*, *HT* is the matching ratio of the hashtag abstract language words including travel-related emotional expressions, and *Vn* represents the emotional probability of *t*.

C. *Emotional word Evaluation*

The emotional direction can be encoded by evaluating positive, neutral, and negative emotional words evaluated in the hashtag of the image set appearing when searching for a specific tourist destination [10]. In the emotion analysis, positivity indicates the positive direction and negative indicates the decrease in the preference in the negative direction. Emotional probability is calculated by extracting the SNS including the tourist destination and then using the term frequency algorithm to find the frequency of reference within the SNS text for each tourism destination. Emotional hash tags refer to tags or keywords that can define emotional probability through collective intelligence. In addition, the emotions expressed in the text are analyzed to reflect positive or negative opinions about the tourist destination. In this process, we use open Hangul API, which is an open platform that can be used in research to identify positive, neutral, and negative emotion words in sufficient number through collective intelligence [11]. The final tourist destination preference index is calculated by using the frequency of comments and emotional analysis results for each tourist destination. By evaluating the combination of hashtags created, it is possible to create a preference index for a specific region and adjust the direction of the related promotion policy.

D. *Evaluation of Emotion Expressions Corpus according to Trend*

The study collected emotional language according to trend through big data analysis and created a database with travel related emotion expression corpus (*HT*). If a match is found in the vocabulary used, such as hashtag searched in the travel destination, it can be evaluated for the purpose of following the trend and rated as preference. Also, it can be used to promote local events, festivals, and specific information in addition to the corpus of the local name through collection of frequent hashtag changed in real time. Figure 5 is part of an example of keyword extraction for travel purposes in the proposed study with purpose specific travel destination classification technique [12].

Figure 5. Emotional language examples

Alone Friendship Date Everyday Healing Friends Sea Vacation Backpacking Guesthouse Daily Family Romance Food porn Communication Weekend Camping Youth Walking Top-restaurant Happiness Man Memories Drive Parenting Itchy-feet Railroad Honeymoon I-love-you Alone Train-travel Nature

It is possible to search common corpus including emotional language that is investigated in real time and maintain it as database (HT) and use it as promotion keyword to improve search efficiency.

IV. CONCLUSION AND FUTURE RESEARCH

This study analyzed the preference of travel destinations by analyzing the hashtag existing in the image set through the collection of SNS, classifying it as emotional language, and comparing it with the vocabulary showing general travel satisfaction. The vocabulary to be evaluated was divided into regional, tourism target vocabulary, common expression language, and emotional expression language. This study demonstrates tourists' preference for travel destinations according to whether collected hashtags match previously retrieved emotional hashtags. Through this, it is possible to analyze the changing travel trend of a specific travel destination and accordingly, it will be possible to establish a travel promotion policy. The limitation of the study is that it is necessary to be able to react to trends quickly because it is necessary establish promotion policies in accordance with real–time trends of recommended emotion language of many existing travel destinations. Through practical data collection and analysis in the future, a study will be done on the conditions and phenomenon of the combination of the hashtag of SNS data and whether they express a strengthening or weakening of promotion effect.

ACKNOWLEDGMENT

This work was supported by a grant from Seoul Accord Fund (IITP-2017-2011-0-00559) of Ministry of Science, ICT and Future Planning (MSIP), IITP

REFERENCES

[1] H. H. Kim, “A Tag-based Music Recommendation Using UniTag Ontology,” The Korean Society Of Computer And Information, Journal of the Korea Society of Computer and Information, vol. 17(11), pp. 133-140, November 2012.

[2] S. H. Lee, S. J. Park, M. H. Lee, and D. H. Hwang, “Efficiency Analysis of Multi Tag-based Web Image Search,” Korea Multimedia Society, pp. 260-263, 2009.

[3] H. J. Park, “A Qualitative Exploration of Folksonomy Users' Tagging Behaviors,” Korean Society for Library and Information Science, Journal of The Korean Society for Library and Information Science , vol. 45(1), pp. 189-210, February 2011.

[4] I. Y. Chae, M. K. Moon, M. S. Kim, and K. Y. Yu, “A proposal on tourist destination preference analysis based on emotion using SNS,” The Korean Society for Geospatial Information Science, pp. 131-132, 2016.

[5] Y. S. Kim, S. H. Jun, M. Jeon, and E. Hwan, “Tourist destination recommendation scheme based on real-time SNS trend analysis”, Korea Information Science Society, pp. 291-293, 2014.

[6] S. J. Kim and S. D. Lee, “A Co-Occuring HashTag Analysis Technique In SNS EnvironMents SNS,” The Korea Contents Society, pp. 223-224, 2014.

[7] Wikipidia, Folksonomy, [Online]. Available from: https://en.wikipedia.org/wiki/Folksonomy/, 2017.07.09.

[8] J. M. Lee, “A Conceptual Access to the Folksonomy and Its Application on the Web Information Services,” The Korean Biblia Society For Library And Information Science, Journal of the Korean Biblia Society for Library and Information Science, vol. 18(2), pp. 141-159, December 2007.

[9] KoNLPy, KoNLPy: Python Korean Langueage NLP, [Online]. Available from: http://konlpy-ko.readthedocs.io/ko/v0.4.3/ 2017.07.09.

[10] J. K. An and H. W. Kim, 2015, “Building a Korean Sentiment L exicon Using Collective Intelligence,” Journal of Intelligence and Information Systems, vol. 21(2), pp. 49-6, 2015.

[11] Open Hangul, Open Hangul API, [Online]. Available from: http://api.openhangul.com/, 2017.05.02

[12] J. H. Moon, J. Rew, Y. Kim, and E. Hwang, “Purpose-based Travel Destination Classification Scheme Using SNS Hashtags,” Korea Information Science Society, pp. 121-123, 2015.

3D Web-Based Shape Modelling: Data Extraction and Delivery

Ali Abdallah
The National Centre for Computer Animation,
Bournemouth University
Poole, United Kingdom
E-mail: ccengineer@gmail.com

Abstract— Despite the rapid development of hardware including specialized graphical processing units (GPUs) and widening the bandwidth, truly interactive applications allowing for near real-time visualization without loss of a visual quality are still to become a reality. Building up an adaptive 3D Web-based shape modelling environment enables us to design platform independent 3D objects in a collaborative manner, yet delivering compressed meshes, and images to clients with different platforms and devices in an optimized manner is still a difficult job. In this paper, we explore the crucial issues of delivering 3D objects. We focus on data transmission over different transmission bandwidths between servers delivering the service and clients with different platform devices. Our case study relies on a client-server adaptive architecture, supported by different rendering techniques, and is able to deliver compressed meshes and images. We identify the obstacles in delivering compressed data files as well as image streams, and present results based on bandwidth capacity, storage size, extraction and loading times.

Keywords- Adaptive architecture; 3D shape modelling; compression; mesh; image streams.

I. INTRODUCTION

Online applications should be efficient and characterized by real time response. Collecting information about clients is a must to keep any online application updated. This process is done using different available scenarios provided by different online and offline resources such as hardware changes, bandwidth monitoring, server stability and performance, memory consumption and availability, etc. [1].

Three major elements constitute the Web -based collaborative 3D shape modelling environment: Networking, Modelling and Rendering. The proposed environment requires a way of communication between its different parts, which is done by the Network element. 3D shape modelling takes place using the Modelling element which constitutes the core of the proposed environment. Rendering is responsible for generating the geometric model after rendering it at the GPU Level.

Polygonal meshes are considered the most common format to store and represent 3D models. They are embodied in the Boundary Representation (BRep). All modern application programming interfaces (APIs) usually support this format, and allow its implementation to the Web browsers. However, the polygonal mesh, by definition, is an approximation of the mathematically precise model that have well-known issues concerned with loss of the precise shape and visual property definition, limited complexity, large memory consumption, problem with transferring through networks, etc. The inability to access the construction history is also an issue. Transmission of 3D scenes is still a major issue in spite of the extensive research dedicated to resolve this kind of problems [2].

Flash animations are widely used on browsers which cannot support large amount of computational resources, and cannot handle large data files that require high Internet bandwidths, power GPUs and large memories. All are accessed by the browser itself. The problem with such option is that it does not support interactivity over the Web. Another option to be mentioned is generating a stream of images of a 3D model, taken from different angles covering 360 degrees, and displaying them on the browser after being arranged and grouped. The object will be loaded as a 3D model to the browser [3].

In a client-server architecture, 3D data is usually generated at the server side, and presented as polygonal meshes which are composed of vertices and facets and can be transmitted over the network [2][4]. Delivering 3D data and visualizing it using a Web browser over the Web is considerably slow and requires a lot of hardware resources (GPU and Memory) [5][6][7].

In this work, we address the problem of extracting and delivering 3D data contents resulted from different rendering techniques. We introduce data extraction methods to support the data delivery process over the client-server architecture; our aim is to optimize the extraction and the delivery process for better performance.

The paper explores different ways of extracting and delivering 3D models based on the used rendering technique. We focus on extracting and delivering 3D models built using a special adaptive 3D shape modeling environment. The paper structure is as follows: after a survey of related works, we describe the adaptive 3D shape modeling environment and identify the types of models generated based on its three rendering techniques. After that, we discuss the data extraction methods used. The paper discusses three different 3D content extractions and delivery methods based on the applied rendering technique, it compares the methods based on bandwidth, data size, computational time and visualization time. Finally, we discuss the results, as well as some practical recommendations reflecting the advantages and drawbacks of the proposed extraction and delivery techniques.

II. Related Work

Modern computers, and even smaller devices such as tablets and smart phones, are equipped with suitable graphic adapters. Kristian et al. focus on improving the global positioning system (GPS) functionality and trying to decrease the expenses resulted from the browser at the client side. They try to increase performance of the current implemented Web features by designing an extension of the available Web document, in the form of 3D scene description format [14].

The huge demands on the transmission of Web -based 3D scenes over the Internet from servers to clients, accompanied by a variety of users with different hardware platforms, devices and GPU powers, prompt researchers focus on mesh compression techniques to ensure the delivery of 3D objects over the Web [2][8][9][12]. Progressive meshes (PM) were presented by Hoppe [10] in 1996, as a cumulative, trimmed and continuous format of polygonal mesh. PM are used in remote visualization, and are capable to display objects' details according to the performance of the requesting client machine [5][10]. In 3D scenes, minimizing the size of the transmitted data by compressing the polygonal mesh is not the only concern. The main objectives of the PM are speeding up the transmission of 3D data, compressing associated attributes like colour and texture, and keeping the quality level of details. Guillaume et al. discuss the above mentioned issues and propose an algorism that allows quick compression for 3D data and generates a binary compressed file [5].

Ramani et al. pointed out that, in order to distribute a model, a special modeller should be installed on each client, so that they can share and modify the same model using the Web browser, in a collaborative manner. They mention that some collaborative modelling systems allow editing some operations such as changing some specific and basic characteristics of the model. One good example of collaboration among users is WebSPIFF [11], which is a tool that allows users to create, modify and delete geometric objects.

Cloud computing empowers different existing technologies such as virtualization and parallel computing by applying some of its characteristics, which include rendering on demand, reliability and efficiency. Cloud computing allows users to benefit from high computing speed over the Web; it also allows storing a large amount of data and real time 3D rendering. Wu et al. described some requirements for their cloud based manufacturing, some of which include the real-time information collection about the clients. Another requirement is 3D models' distribution and sharing by accessing big data files stored on the cloud, as well as processing and managing these data files for modelling and rendering purposes [22].

One of the 3D data compression methods is the curvature prediction method, discussed by Adrien et al. This method is supported with a wavelet formulation method designed to improve rate-distortion (R-D) performance, it is a quantization method to increase the compression rate. These methods are applied to a simple algorithm designed for 2-manifold mesh compression [9].

Xinshu et al. discuss the problem of 3D sensitive information, and suggest an intra-origin data control system (CRYPTON), that allows owners to monitor and direct sensitive 3D data loaded to the browser at the client side [13].

III. Adaptive 3D Shape Modeling Architecture

The used modelling environment, which is an Adaptive 3D shape modelling architecture (ASMAR), is based on Hybrid representation or HRep, which is the integration and combination of both function and boundary representations. HRep acquires the characteristics of both functional representation (FRep) and boundary representation (BRep), and allows dealing with objects as volumes having an internal structure; it keeps the constructive tree of the modelling process, and allows the representation of the model in the form of a polygonal mesh.

BRep represents 3D objects in the form of a polygonal mesh, which is the approximation of the mathematically precise model. 3-dimensional objects are stored and represented as a set of inter-related facets and vertices, which when rendered, generates an image of a 3D model. This representation can be implemented in Web browsers because it is widely supported by the modern 3D APIs [15]. BRep can be stored in different 3D file formats using the polygonal mesh. Such files can hold all the necessary data needed to render the 3D object at the GPU level and they are easy to be created, stored, and accessed. One of the main concerns of such representation is about the loss of the model precision with the fact that BRep, as mentioned before, is an approximation of the real object. Another concern is about the lack of construction history of the object. Whenever the object is subjected to a change, the original 3D object will be lost and replaced by the latest version of the edited object. Another drawback is that BRep cannot handle objects with internal structures, and it is only concerned with the external surfaces of 3D objects rather than their internal structures. The size of 3D data files can be enormous when dealing with complex objects and facets. Serious problems can arise when transferring the files over the Internet, in addition to large memory consumption.

FRep is based on function representation of 3D models rather than facets and vertices. Such representation reduces the complexity of 3D models and can easily manipulate complex objects without the need of handling large data files and a huge amount of computational resources. HyperFun, which is an open source and high level programming language, uses FRep [16] to implement complex 3D models; HyperFun files are of small size and can handle complex geometric objects [16]. Even though FRep reserves the development track of the model, there are still major weaknesses in using FRep models, such as the fact that they are difficult to be treated and controlled inside a Web browser. Browsers load 3D models using polygonal meshes. This allows users to edit and modify 3D scenes and objects in an interactive manner [18]. One of the serious problems when using FRep, is the need for Java Applets to generate and display HyperFun objects to the Web browser, which could be a very expensive procedure and need vast amount of computational resources.

HRep, as mentioned before, is the combination and integration of both FRep and BRep modeling techniques. It combines the two and synchronizes them, trying to make use of their characteristics in an efficient way, by eliminating the drawbacks of each one.

Since HRep is composed of two major components, the conversion to these two components is a must. The conversion of FRep to BRep can be done through a process called Polygonization, which is a way of converting FRep object surface into a polygonal mesh by creating flat polygons to approximate shapes. Another way of conversion FRep to BRep is voxelization, which is a procedure responsible for the extraction of voxel representation from a continuous geometric representation. The extraction process can be very expensive in terms of time and resources [19]. The bi-directional conversion process, in Fig. 1, represents the conversion of BRep to FRep, using signed distance fields. Starting from a given polygonal mesh, using the distance-fields technique, we are able to obtain the exact function representation for that mesh, both the interior and the exterior of the object can be represented [20].

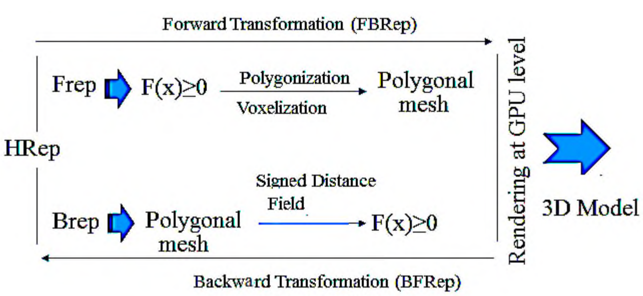

Figure 1. HRep bi-directional conversion showing a two way transformation between FRep and BRep (forward-backward).

IV. 3D DATA EXTRACTION AND DELIVERY

A. 3D Model Data File

3D models are often stored as a set of 3D points called vertices. Indexing these vertices, allows to construct polygons, which constitute the surface of the 3D model [21].

Object (Obj) file is based on text file with ".obj" extension, it is an extensible 3D graphics (X3D) text based. It allows user authorship, and it is considered as an environment for 3D scenes construction using XML or classic VRML encoding [21]. Open file format was developed by Wavefront [15]. 3D Obj file was adopted by different 3D graphic vendors because it was easy to be imported and exported. Table I shows the structure of the file. The indicator is a single command line followed by a series of values representing the indicator. This digital data is loaded into the Obj file. Our 3D shape modeling environment is supported by special tools that allow to access the GPU buffer, and to extract the rendered 3D object data as vertices and facets and to save them to the Obj file [15].

TABLE I. OBJ FILE INDICATORS AND VALUES DESCRIPTION

Indicator	Description
#	Comment
v	Vertex
vn	Normal
vt	Texture Coordinate
f	Face
l	Line
G	Group

Extracting the 3D model from the GPU buffer takes place using the Marching Cubes Algorithm [15], which is one of the most applied rendering techniques used in 3D modeling. The extraction process allows to access the GPU memory and to extract both the vertices and the facets .The extracted vertices and facets are re-assembled where suitable obj indicators are added before being saved into the text based Obj file.

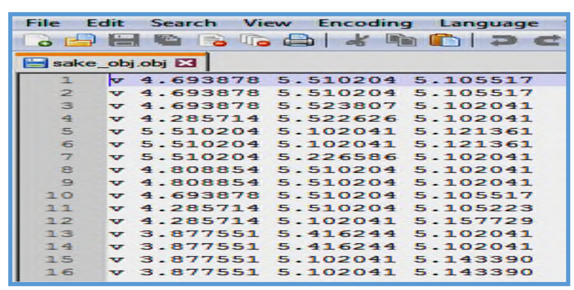

Figure 2. Obj File showing indicator v as vertex and three different extracted points next to each vertex as vertex values.

When completed, the obtained Obj file is loaded using OBJLoader, a special JavaScript function declared in Three.js JavaScript file engine. The OBJLoader function is responsible for loading 3D models with Wave front file format using Three.js as shown in Fig. 3 [23]. The code below shows X3D file embedded in XML encoding.

```
<X3D id="x3d">
 <Scene id="sc1">
   <Viewpoint position="0.15255 0.10231 0.19884"
```

```
            orientation="- 0.27505 0.95696 0.09264
            0.55979" >
    </Viewpoint>
    <Transform id="Hemi">
      <Inline url="Hemi-spher.x3d" solid="false"></Inline>
     </Transform>
  </Scene>
</X3D>
```

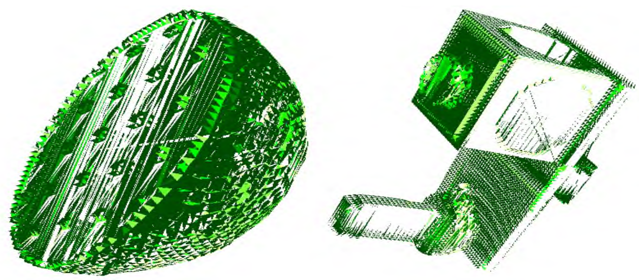

Figure 3. Hemi-sphere and sake-pot complex objects loaded from Obj files.

B. Data Extraction and Delivery

The data extraction and delivery process passes through different phases where different extraction and delivery techniques will be applied according to the technique used in rendering the 3D model. As shown in Fig 4, the Marching Cubes and Hybrid WebGL use the same extraction technique by accessing the GPU memory, extracting all vertices and facets, saving them to a predefined Obj file and loading the Obj file into the browser.

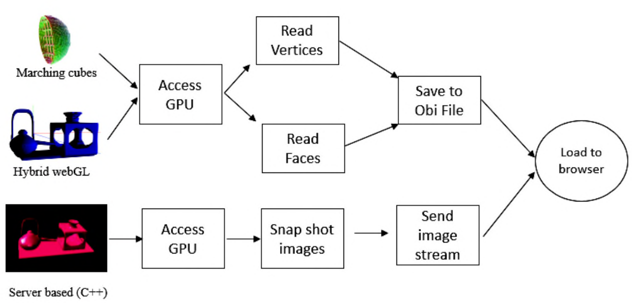

Figure 4. Work flow model showing three different data extraction and delivery processes applied on three different rendered techniques starting from GPU level.

Server based rendering uses a different approach. It starts by accessing the GPU, not to read the points from memory, but to take snapshots for images, from different angels, then it sends the images as an image stream to the client after saving them in an image matrix. The image matrix is then displayed in the browser, as indicated in Fig 4.

The data extraction process is designed to extract and deliver 3D data safely over the Internet. Its role is to access the GPU buffer, extract all the needed data, save them in Wavefront (Obj) file format, then deliver the Obj file over the transmission media. Using (1), we can obtain all points that constitute the 3D model [23].

$$pt = \sum_{z=0}^{s} \sum_{y=0}^{s} \sum_{x=0}^{s} x+s*y+s^2*z. \qquad (1)$$

Where
s: Size of the geometric cube.
pt : 3D point.
vert[]: array of extracted 3D point vertices.
vindex: vertex index.
va[] : array of all vertices.
facet: array of facets .

$$ptx = pt+1 \qquad (2)$$
$$pty = pt + s \qquad (3)$$
$$ptxy = pty + 1 \qquad (4)$$
$$ptz = pt + z^2 \qquad (5)$$
$$ptxz = ptx + s^2 \qquad (6)$$
$$ptyz = pty + s^2 \qquad (7)$$
$$ptxyz = ptxy + s^2 \qquad (8)$$

The process of extracting a vertex is shown in (9)

$$vert[] = \sum_{i=0}^{3} \text{vertices[vindex][i]} \qquad (9)$$

The vertices of the whole object can be obtained using (10)

$$va[] = \sum_{a=0}^{Nb} vert[a] \qquad (10)$$

And the facets can be obtained using (11).

$$face = \sum_{a=0}^{Nb} \sum_{i=0}^{3} face[a][i] \qquad (11)$$

In our study, since we have three types of rendering techniques (Marching cubes, Hybrid WebGL and Server based using C++), it is useful to do experiments on the three rendering types. Our experiments are applied to three different models, one simple model (Android) and two complex models (Hemi-sphere and Sake pot).

All experiments run over a 64-bit windows 7 operating system, use Intel core i7-6700 CPU at 3.4 GHz. The available installed RAM is 16 GB, the GPU being used is Intel HD Graphics 530 (GT2) with 1150 MHz clock with ability to access the main memory (2x64 bit DDR3L-1600), we use google chrome browser version 59.0.3071.115 (64 bit).

Since we are using dynamic complex objects that are subjected to different parameter changes, the 3D data extraction process will directly be affected by these changes. One of the parameters is the resolution parameter. In our experiment, we apply the extraction process to the resolution test, in order to monitor the extraction process time and size. The extraction test of the resolution parameter is performed on the three different models we have.

C. Marching Cubes Data Extraction and Delivery

In this study, we use three objects, Hemi-Sphere and Sake Pot (both are complex objects) and Android robot (simple object). The loaded files formats are OBJ and X3D file formats.

TABLE II. RENDERING USING MARCHING CUBES

Rendering using Marching Cubes			
	Hemi-Sphere (complex object)	Sake Pot (complex object)	android (simple object)
Loading time in sec	0.382 sec	0.275 sec	0.104 sec
No. of Vertices	52504	55652	2828
No. of Facets	27119	27814	1352
Total	79623	83466	4180
Loading Obj File in sec	0.559 sec	0.598 sec	0.105 sec
OBJ File Size in KB	1707 KB	1875 KB	86 KB
Loading X3D File	0.001 sec	0.002 sec	0.001 sec
X3D File Size in KB	1976 KB	1902 KB	72 KB

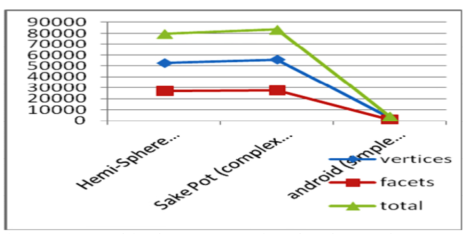

Figure 5. Graphical representation for the total extracted vertices and facets, comparing the 3 different objects.

After the extraction process is performed, we obtain the following results, as shown in table II.

- The loading time difference between the three objects is minor, with no big difference.
- The extracted facets and vertices for the two complex objects are approximately the same, and they are double of those of the simple objects
- The OBJ loading time of all objects is acceptable and is less than 0.6 sec.
- The file size of the complex objects is relatively big (more than 1.5 MByte) while that of the simple objects is relatively small, less than 100 Kbytes.
- The loading time of the X3D file is significantly better than that of the Obj file and is considered relatively short.
- The X3D file size is close to the Obj File size of the three objects.

Table II shows the variation in loading extraction time between 0.104 sec for simple object and 0.382 sec as max for the Hemi-sphere complex object. This variation indicates that the extraction process can be very fast and with no major difference between simple and complex objects when applying Marching Cubes rendering. We can notice that the total number of extractions for Hemi-sphere (79623) is less than that for the sake-pot (83466) by 5% while the extraction time for Hemi-sphere is considerably higher (0.382 sec) than that of the sake-pot (0.275 sec) by almost 37%. The fact is that the extraction time is not directly proportional with the extracted number of facets and vertices, but it is related to the complexity level of the rendered object itself. In our case, the Hemi-sphere model is considered more complex than the sake-pot model and that is the reason why the extraction process taking more time. The number of the extracted facets and vertices in Fig. 5 shows that the more vertices and facets to be extracted, the longer the extraction time.

D. Hybrid Data Extraction and Delivery

This test is applied to two complex objects using the Hybrid rendering technique. Here, the 3D complex objects are subjected to a resolution parameter. The extraction process takes place in six different phases, where the resolution parameter is changing accordingly.

TABLE III. EXPERIMENT APPLIED ON THREE DIFFERENT OBJECTS USING HYBRID RENDERING.

Density	Load time in sec	No. of Vertices	Vertices Extraction Time	No. of Facets	Facets Extraction Time	Total extraction Time	Obj Size in KB
Hemisphere: Complex Object using Hybrid modeling							
30	0.1 sec	12868	25.7 sec	12868	17.2 sec	42.9 sec	769 KB
40	0.4 sec	12868	60.0 sec	12868	73.3 sec	133.3 sec	1787 KB
50	0.4 sec	38160	419.9 sec	38160	285.7 sec	705.7 sec	2341 KB
60	0.9 sec	55648	1076.8 sec	55648	918.5 sec	1995.4 sec	2751 KB
70	1.5 sec	73568	3584.2 sec	73568	2544.6 sec	2128.8 sec	4666 KB
80	1.7 sec	94584	8544.0 sec	94584	6585 sec	4529.0 sec	5227 KB
Sake-Pot: Complex Object using Hybrid modeling							
30	0.1 sec	2636	5.7 sec	2636	3 sec	8.6 sec	157 KB
40	0.2 sec	7848	39.0 sec	7848	21.5 sec	60.5 sec	322 KB
50	0.2 sec	16460	180.2 sec	16460	90.0 sec	270.2 sec	531 KB
60	0.5 sec	28468	136.8 sec	28468	121.6 sec	258.4 sec	699 KB
70	1.3 sec	37492	124.8 sec	37492	280.2 sec	405.0 sec	1053 KB
80	2.2 sec	40488	1140.2 sec	40488	1140.2 sec	1080.4 sec	1355 KB

From the results shown in Table III, we conclude the following:

- The loading time is increasingly changing with the changing of the resolution of the extracted objects. The higher the resolution, the longer the loading time needed to display the object. A simple comparison shows that loading an object of 30% resolution takes a small fraction of a second (0.179 sec for a Hemi-sphere) while loading the same objects of 80% resolution takes 1.701 seconds, which means 10 times more time is needed to load the same object.

- The number of extracted points (vertices and facets) increases as the resolution of the extracted object increases. Therefore, the file size increases.
- Table III shows that the more complex the 3D object are, the longer extraction time is needed, because more facets and vertices are to be extracted. Therefore, more storage space is needed to save the Obj files.
- Fig. 6 shows a large difference in the extraction times between two different complex objects. The extraction time is directly affected by the resolution of the object, the level of complexity and the number of extracted facets and vertices.
- Fig. 7 compares the Obj file sizes, and shows that the higher the level of complexity, the more storage is needed to save the Obj files. In our experiment, the Hemi-sphere needs more than 5 Mbytes to store the Obj file of 80% resolution while it needs less than 1.5Mbyte to store the Obj file for the sake-pot model at the same level of resolution.

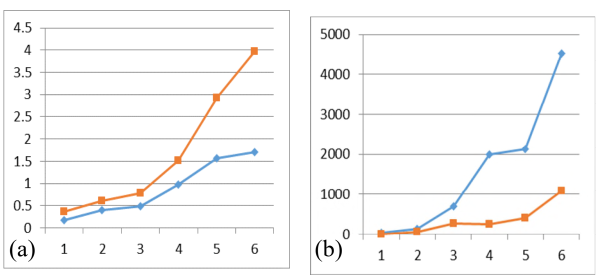

Figure 6. Graphical representation for the time taken (a) and the extracted values (b) for Hemi-sphere (orange line) and Sake Pot (blue line) respectively when applying the resolution parameter.

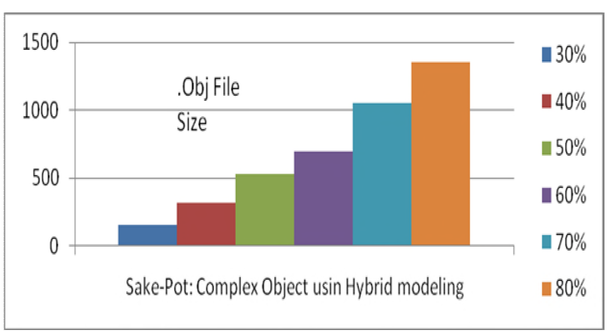

Figure 7. Bar chart representation illustrating the Obj file size with respect to resolution parameter. As the resolution parameters increase, the file size increases.

E. Server Based Data Extraction and Delivery

TABLE IV. EXPERIMENT APPLIED ON SAKE-POT OBJECTS USING SERVER BASED RENDERING

Snap-shot rate	Matrix size	No. of images	Time taken	Images/sec	Image size\Kbytes	Total size
2	45*45	2025	198 sec	10 image/sec	100 KB	197.7 KB
3	42*42	1764	224 sec	7.8 image/sec	100 KB	172.2 KB
4	38*38	1444	274 sec	5.2 image/sec	100 KB	141.0 KB
5	35*35	1125	160 sec	7.0 image/sec	100 KB	109.8 KB
6	32*32	1024	166 sec	6.1 image/sec	100 KB	100 KB
7	28*28	784	126 sec	6.2 image/sec	100 KB	76.5 KB
8	25*25	625	112 sec	5.5 image/sec	100 KB	61.0 KB
9	22*22	484	104 sec	4.6 image/sec	100 KB	47.2 KB

Server based rendering generates images with fixed size; the precision of the object depends on the number of generated images. The snap-shot rate determines the rate of captured images per sec. In our test, we apply eight different rates starting from two and increasing up to nine. As the rate increases, the number of generated images decreases.

Table IV shows that at rate two, the image matrix is 45 by 45 or 2025 images, while at rate nine, the image matrix becomes 22 by 22 or 484 images only. As the snap-shot rate increases, the number of captured images and the time taken to generate the images decreases; therefore, the total size of the generated images decreases as well. That means, the higher the snap-shot rate, the more precision we obtain, and more storage size and image generation time are needed.

Table IV reveals that when the image matrix starts to shrink, the number of images starts to decrease, and the time taken to generate the images starts to decrease too, since the image size is fixed, the storage size needed to save all the generated images is directly proportional to the number of generated images. As a result, when the image matrix shrinks, the precision of the 3D model is reduced, and the storage space needed to store the object images is reduced as well.

V. CONCLUSION

In this work, we identify and implement a special extraction and delivery data module to extract the online mesh (vertices and facets) and deliver the raw data after being loaded into different file formats or image streams.

The visual and numerical results have lead us to the following conclusions:

- Extracting raw data from the GPU buffer can be done. The extraction time depends on the level of

Copyright (c) IARIA, 2017. ISBN: 978-1-61208-587-6

complexity of the object rather than the number of facets and vertices that exist.

- Raw data extracted from Marching Cubes or hybrid rendering is delivered using Obj and X3D file formats.
- Extracted objects using server based rendering use image stream to deliver the 3D object.
- Loading the extracted file (.obj) which hides the functions and data from users, and displays 3D models on the browser can be done.
- There is a considerable difference when extracting raw data between simple and complex objects, in terms of extraction time, file size and loading time.
- Image streams can be a good solution to clients with low GPU power, but require more storage and bandwidth.

When comparing the three extraction tests, we can conclude the following:

- The loading time for Hemi-sphere with Marching Cubes rendering is almost equivalent to that of Hybrid rendering at 40% resolution.
- The loading time for Hybrid rendering with 80% resolution is 4.4 times higher than that of marching cubes. That is why Hybrid rendering is best used with powerful GPUs.
- The Obj file size loaded with Hybrid rendering at 80% resolution (5227 KB) is considerably smaller compared with that loaded from Marching Cubes rendering (1707 KB). That means we need more extraction time and more GPU power in Hybrid rendering, but lower storage to save Obj files compared to Marching Cubes rendering.

REFERENCES

[1] Y. Jung, J. Behr, T. Drevensek, and S. Wagner, “Declarative 3D Approaches for Distributed Web-Based Scientific Visualization Services,” Proceedings of the 1st International Workshop on Declarative 3D for the Web Architecture, 2012.

[2] M. Limper, S. Wanger, C. Stein, Y. Jung, A. Strok, F. IGD, and T. Darmstadt, “Fast Delivery of 3D Web Content: A case Study, Web3D,” Proceedings of the 18th International Conference on 3D Web Technology, pp. 11-17, 2013.

[3] M. Rodrigues, M. Kormann, and L. Davison, “A Case Study of 3D Technologies in Higher Education: Scanning the Metalwork Collection of Museums Sheffield and its Implications to Teaching and Learning,” International Conference on Information Technology. IEEE, 1-6, 2011.

[4] A. Evans, M. Romeo, A. Bahrehmand, J. Age, and J. Blat, “3D Graphics on the Web: a Survey,” Interactive Technology Group, University Pompeu Fabra, Barcelona, Spain. 2014.

[5] H. Lee, G. Lavoue, and F. Dupont, “Rate-distortion Optimization for Progressive Compression of 3D Mesh with Colour Attributes,” The Visual Computer, 137-153, 2012.

[6] M. Rezayat, “The Enterprise-Web Portal for Life-cycle Support,” J. Computer-aided design, 2000: 32 .85-96.

[7] S. Abrahamson, D. Wallace, N. Senin, and P. Sferro, “Integrated Design in a Servicemarket Place,” Computer-Aided Design, 2000.

[8] J. Peng, C. Kim, And C. Jay Kuo, “Technologies for 3D Mesh Compression: A survey,” Journal of Visual Communication and Image Representation. Volume 16, issue 6, 2005.

[9] A. Maglo, C. Courbet, C. Alliez., and C. Hudelot, “Prouressive Compression Of Manithld Polyzon Meshes,” Computers and Graphics, Shape Modeling International (S1\11) pp. 349-359. 2012.

[10] H. Hoppe, “Progressive Meshes,” Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 99-108. 1996.

[11] K. Ramani, A. Agrawat, and M. Babu, “CADDAC: Multi-Client Collaborative Shape Design System with Server-based Geometry Kernel,” in Journal of Computing and Information Science in Engineering, volume 3, issue 2, 2003.

[12] H. Grasberger, P. Shirazian, B. Wyvill, and S. Greenberg, “A Dataefficient Collaborative Modelling Method Using Websockets and the Blobtree for Over-the Air Networks,” in Proceedings of the 18th International Conference on 3D Web Technology. New York, pp. 29–37, 2013.

[13] X. Bong, Z. Chen, H. Siadati. S. Tople. Saxena, and Z. Liang, “Protecting Sensitive Web Content From client-side vulnerabilities with CRYPTONs,” Proceedings of the 2013 ACM S1GSAC Conference on Computer & Communications Security pp. 1311-1324, 2013.

[14] K. Sons, F. Klein, D. Rubinstein, S. Byelozyorov, and P. Slusallek, “ XML3D- Interactive 3D graphics for the Web,” Proceedings of the 15th International Conference on 3D Technology, pp. 175-184, 2010.

[15] A. Abdallah, O. Fryazinov, V. Adzhiev, and A. Pasko, “3D Web-Based Shape Modelling: Building up an Adaptive Architecture,” ACI-11 2014 . The Seventh International Conference on Advances in Computer-Human Interactions. 2014.

[16] R. Cartwright, V. Adzhiev, A. Pasko, Y. Goto, and T. Kunii, “Web-based Shape Modelling with Hyper-Fun,” ILIN Computer Graphics and Applications. Vol. 25. pp. 60-69, 2005.

[17] C. Vilbrandt, G. Pasko, A. Pasko, P. Fayolle, T. Vilbrandt, and J. Goodwin, “Cultural Heritage Preservation Using Constructive Shape Modeling,” Volume 23, number 1 pp. 25-41, 2004.

[18] O. Fryazinov, A. Pasko, and V. Adzhiev, “An Exact Representation of Polygonal Objects by Differentiable Scalar Fields Based on Binary Space Partitioning,” Technical Report “TR-NCCA-2008-03”, The National Centre for Computer Animation. Bournemouth University. UK. 2008.

[19] J. Pantaleoni, “VoxelPipe: A Programmable Pipeline for 3D Voxelization,” (NVIDIA), in High Performance Graphics, 2011.

[20] M. Sanchez, O. Fryazinov, and A. Pasko, “Efficient Evaluation of Continuous Signed Distance to a Polygonal Mesh,” 12 Proceedings of the 28th Spring Conference on Computer Graphics, pp. 101-108, NY, USA, 2012.

[21] K. McHenry, and P. Bajcsy, “An overview of 3D Data Content File Formats and Viewers,” National Center for Supercomputing Application. USA. 2008.

[22] D. Wu, D. Rosen, L. Wang, and D. Schaefer, “Cloud-Based Manufacturing: Old Wine in New Bottles? Variety Management in Manufacturing,” Proceedings of the 47th CIRP Conference on Manufacturing Systems, 2014.

[23] A. Abdallah, “Securing Online 3D Web-Based Models”, ICT and Societal Challenges, LAU Beirut / New York, 2017.

Methodology of Analysis of Users' Information Demand through Search Keyword Analysis

Eunji Yu / KISTI
NTIS Center
Daejeon, Korea
eunjiyu08@kisti.re.kr

Wonkyun Joo / KISTI
NTIS Center
Daejeon, Korea
joo@kisti.re.kr

***Abstract*—The search records in a portal service which provides diverse information services reflect users' demand for information and services. Therefore, the analysis of search records in the portal service makes it possible to figure out and measure the demand for user-wanted information or services. This study suggests a model which investigates users' information demand for national Research and Development (R&D) data after analyzing the National Science & Technology Information Service (NTIS) users' search records. The NTIS has collected national R&D data such as government R&D programs, projects, research outcomes and equipment. It appears that that users' information demand measured through the methodology in this study would be helpful in developing service strategies and establishing services.**

Keywords-Government R&D; Text Analysis; Topic Modeling; Information Demand.

I. INTRODUCTION

A portal is a service designed to provide diverse information services to allow users to search wanted information on a certain site without visiting every website. Recently, the portal service has attempted to provide user-customized services after improving the conventional information search-oriented service. For the effective promotion of this mission, it is critical to figure out what information and services users want. For this, one of the most commonly used methods is to analyze the keywords of the users who use the portal. It appears that users' keywords are the data reflecting their demand for information and services in a realistic manner. Therefore, keywords have been deemed reasonable data needed to figure out users' demand. In fact, there have been a lot of efforts to improve information services based on the analyzed results in diverse fields. This study aimed to analyze the keywords of the NTIS (National Science & Technology Information Service) [1], which has collected and provided national R&D information, such as national R&D programs, projects, research outcomes and equipment. In addition, the study investigates users' demand and its trend with a goal of proposing a model which can analyze how users' information demand is changing.

The rest of the paper is structured as follows. In Section II, conventional studies on the research and keyword analysis methods designed to measure users' information demand are briefly mentioned. In Section III, keywords are analyzed and a scheme to construct a model designed to figure out users' information demand is explained. Lastly, the contribution, limitations of this study and future research directions are stated.

II. RELATED RESEARCH

2.1. NTIS Service

NTIS provides various services for researchers, decision makers, and governmental officers by converting national R&D information such as programs, projects, human resources, research facilities and equipment into information in a database by the interconnection with 17 governmental departments conducting national R&D projects for improving the effectiveness of research and development through the research life cycle from the R&D planning and the utilization of the outcomes [2].

2.2. Demand Measurement

With the development of information technology, it has become more important to predict users' information demand to properly respond to the fast changing trends. To figure out the exact demand, a lot of resources are needed, such as optimum data, which contains users' demands and domain experts' knowledge. Based on these resources, in fact, a variety of demand prediction techniques have been used. They are divided into quantitative and qualitative techniques. One of the most popular qualitative techniques is a questionnaire survey using experts' opinions. With high accuracy, this method is widely used among businesses. However, with this technique, it is difficult to define a scope of experts or consumers. In terms of quantitative techniques, there are regression analysis, time series analysis and diffusion model-based demand prediction method [3]. When the demand is predicted using a quantitative technique, it is very important to decide what data would be analyzed. The technique is used when optimum data can be secured. During time series analysis, in particular, it is usually adopted in the fields where data can be easily collected. Furthermore, this method is able to model complicated causality, such as diverse variables and time difference, and predict the future by discovering the flow of past data and new trends.

2.3. Keyword Analysis Service

Lately, portal services have provided a service which measures users' interest with the accumulated keywords and analyzes trends. For example, Google has introduced Google Trend [4], which shows keyword query count in an index format based on the keywords of the users from around the world. Naver [5], the largest portal site in the Republic of Korea, has also provided a keyword statistics service called 'Naver Trend'. Since the nation's search engine market share is dominated by Naver (over 70%), it is appropriate to measure the domestic interest on keywords [6]. The predictions by Google or Naver trend analysis have little difference, even compared to reliable statistical data. To examine users' demand through analysis of keywords, therefore, there should be studies on the subjects which meet the conditions further specified. The information to be analyzed should be representative or comprehensive or should have many users, like Google and Naver. Therefore, in this study, the search data of NTIS which has the Korean government R&D information was selected as the analysis target.

III. ANALYSIS MODEL FOR USERS' INFORMATION DEMAND

A model proposed in this study is shown in Figure 1. The topic analysis methodology used in the model in this study derives a set of the keywords extracted by calculating the importance of the terms using probability based on the TF-IDF (Term Frequency-Inverse Document Frequency) [7] values. The topic analysis differs from conventional models in that it can prevent similar or the same keywords from being classified into a different category when keywords are analyzed based on the TF only. Lastly, the importance of the keywords derived through the topic analysis is visualized, using probability values, as shown in Figure 3.

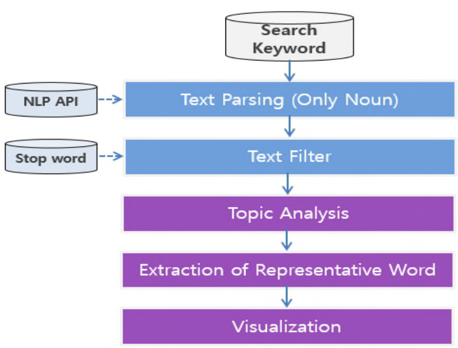

Figure 1. Analysis model for users' information demand

In Figure 1, the process can be explained as follows: In Step 1, pre-treatment on the keywords about national R&D information in the NTIS for the past year is performed. Nouns are only extracted by reflecting users' search query properties, and homonyms are processed using a synonym dictionary. Furthermore, unused terms are filtered with a stopword dictionary. Since the unstructured data 'texts' are analyzed, a pre-treatment process which handles homonyms and disuse is very important in this methodology, even having influence on final analysis results. In Step 3, topics are analyzed. The result of topic analysis is shown in Figure 2.

Topic_id	Doc_Cutoff	Term_Cutoff	Topic	Ic_Numterms	Numdocs
1	0.002	0.003	Korea Pharmacy, pancreatitis, molecule, diagnosis, atopy	56	2156
2	0.002	0.003	High land, cultivation	406	3538
3	0.002	0.003	small, ship, diesel, engine, sunlight	477	12689
4	0.002	0.003	Online, e-commerce	518	13407
5	0.002	0.003	electricity, power, system, storage, technology	300	4558
6	0.002	0.003	High-definition, wireless, image	536	6215
7	0.002	0.003	Pump, material, Antifouling, impurities	516	8696
8	0.002	0.003	disaster, map, risk	519	9614
9	0.002	0.003	stage setting, spray, helmet	510	7694
10	0.001	0.003	fourth Industrial Revolution, science technology, promotion	172	5356

Figure 2. Result of topic analysis

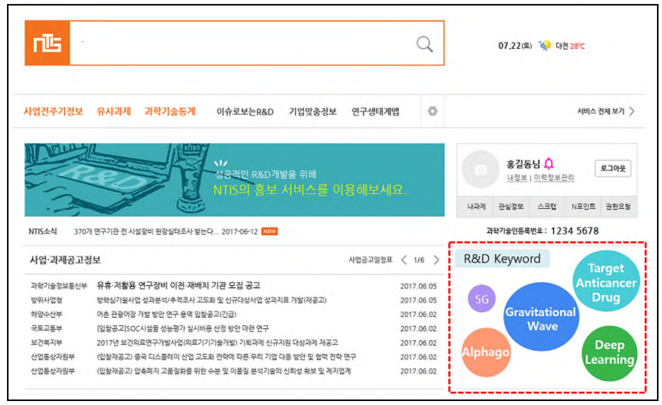

Figure 3. Visualization example of analysis result

As a result, we are able to find out users' demand for national R&D information and demand trends at a glance, as shown in Figure 3.

IV. CONCLUSIONS

This study has proposed a model which can measure users' demand by analyzing the keywords in the NTIS and even figure out the demand trends on national R&D information. This research is significant in that it suggested methodology through which users' demand for national R&D information, which is relatively difficult in terms of demand prediction and its changes, can be analyzed as a field in which a huge amount of information is stored, diverse topics are mixed, and new research topics are continuously produced. In future studies, however, it is necessary to verify this methodology by analyzing actual NTIS keywords. In addition, keywords are simpler than conventional studies in which texts are analyzed in terms of pre-treatment. To derive filtered results, however, a high-quality glossary and a stopword dictionary in national R&D information must be developed in advance.

ACKNOWLEDGMENT

This research was supported by Construction of NTIS funded by the Ministry of Science, ICT and Future Planning.

REFERENCES

[1] National Science & Technology Information Service, http://www.ntis.go.kr [accessed July 2017]

[2] MyeongSeok Yang et al., "A study on the NTIS service," Journal of Korea Technology Innovation Society, 13 2013, pp.294-304.

[3] "Methodology on Short-Term Demand Forecasting: ARIMA model," 34, (2004), pp.68-80.

[4] Goodle Trends, https://trends.google.com/trends/ [accessed July 2017]

[5] Naver Trends, http://datalab.naver.com/keyword/trendSearch.naver [accessed July 2017]

[6] B. Cogille, J. Wolfers and E. Zitzewitz, "Using Prediction Markets to Trak Information Floes: Evidence from Google," American Economic Association, 2009, pp.1-44.

[7] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann Pulishers, San Francisco, 2011.