Measuring the Impact of Content Adaptation on Social Media Engagement

Zaenal Akbar, Ioan Toma, Jose Maria Garcia, Dieter Fensel
Semantic Technology Institute, University of Innsbruck, Austria
zaenal.akbar@sti2.at
ioan.toma@sti2.at
jose.garcia@sti2.at
dieter.fensel@sti2.at

Abstract: The ultimate goal of social media marketing is to reach and engage with the widest audience possible. Usually, the content published on social media channels must be adapted in order to meet the requirements of the target channels. This process, known as content adaptation, requires human knowledge to determine how the content should be prepared to meet the channel requirements, and to spread it across the channel as fast as possible. Extracting this knowledge for content adaptation is a challenging task due to the constant evolution of the variety of human experiences and social media channels. In this paper, we show how we acquire and learn this knowledge as our first step towards content adaptation automation. We determine the social media posts properties that could be affected by the content adaptation. Those properties are: (i) presentation property where a post can be seen as a collection of terms, (ii) hypermedia property where a post can be seen as a collection of media connected by hyperlinks, and (iii) named-entity property where a post can be seen as a collection of real world entities. Based on these properties, the most engaged posts are selected and their similarity degrees to their original sources are computed and compared. From the similarity degree comparison, we can devise recommendations on how to obtain the highest engagement on a particular social channel by fine-tuning the content adaptation on each property.

Keywords: content adaptation, social media, engagement, automation

1. Introduction

Social media introduces various ways of communicating and connects to a wide audience. Being able to reach and engage with the widest audience possible is the ultimate goal of online marketing. With the constantly growing number of communication channels, being present on multiple channels is only the first step; the next step is to be able to deliver high quality content to the audience on each channel. This content dissemination to those multi online communication channels can be achieved through:

1. *Individual dissemination*. Content is disseminated into each channel by users individually.
2. *Mass dissemination*. Content is entered to various channels simultaneously, typically by using a content dissemination tool that could disseminate content to multi-channel at once.
3. *Content sharing dissemination*. Content is entered to a communication channel, and then shared the content to other channels.

Since each channel possesses unique characteristics and requirements, a process of fitting content to multiple channels is an important necessity. This process is known as “content adaptation”: a process of adapting content in ways which the content can then be entered into the relevant channel. More than that, an adaptation could also be affected by the intention of the communication. The process is mainly determined and performed by human users. Dissemination tools contribute only marginally to the process, mainly in the technical aspect of dissemination, for example, ensuring the content types and the length of the content is accepted by the channels. Nevertheless, the main processes of dissemination are still handled by humans, for example, in the determination of which content should go to which channel and how this should be disseminated.

It is commonly used to measure the effectiveness of content dissemination on social media based on user engagement, for example: likes, dislikes, number of comments, and so on. To the best of our knowledge, there is no measurement yet for the dissemination effectiveness based on the impact of the adaptation process to an input content. Being able to measure the impact of content adaptation could give us a better understanding on how to perform the adaptation in order to obtain a better engagement.

In this paper, we would like to evaluate the impact of content adaptation to the user engagement on social media. Our evaluation is based on the measurement of similarities degrees between original and adapted content to the obtained engagement measures when publishing the adapted content to a social media channel. The main contributions of this work are to:

i. define three different content representations which could be affected by an adaptation,
ii. determine the most engaged contents based on these representations,
iii. measure and compare the similarity degrees of original and adapted contents on multiple channels.

In chapter 2 we introduce the content adaptation in online communication and how to measure its adaptation impact. Chapter 3 presents our evaluation results on social media channels from tourism industries, followed by related work in chapter 4 and our conclusion and future work in chapter 5.

2. Content Adaptation Impact Measurement

In this section, we describe and formulate the impact of content adaptation to the user engagement in online communication, especially social media. First, we describe the scope of content adaptation, followed by the content representations which could be affected by the adaptation, and finally we describe how to measure the impact of this adaptation based on the user engagement measures.

In this work, we define a content adaptation as a process of adapting an input content according to the given attributes of an individual online communication channel, where the main objective is to "fit" the content to the channel such that the content can be published on the channel and reach the widest audience possible. The adaptation process is affected by the attributes of a channel (i.e. channel requirements such as the types of content accepted) as well as the publication intention (i.e. as reminder).

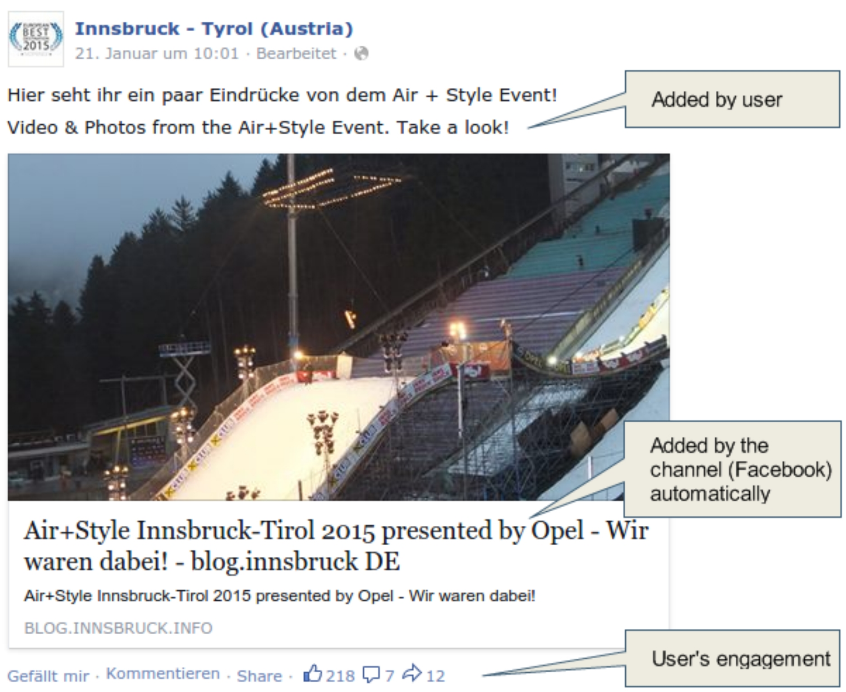

Figure 1: A content adaptation example (source: https://www.facebook.com/Innsbruck)

Figure 1 illustrates an example of content adaptation where a webpage was shared on the channel Facebook. There, two adaptations were applied; first, by the user who shared the page by adding additional information, and second from the channel itself (in this case Facebook) who extract an image, the title and description of the page and include them in the post. The picture also shows the user engagement measures for that post.

In order to perform the adaptation in multi-channel communication, an adaptation user (based on his/her knowledge), must determine the input content, specifications or requirements of targeted channels. He/she then performs the content adaptation to produce output contents which are suitable to be published to each channel, as shown in Figure 2. As shown in this picture, the user must carefully and manually adapt an input content to fit into each channel due to the variety requirements of the channels.

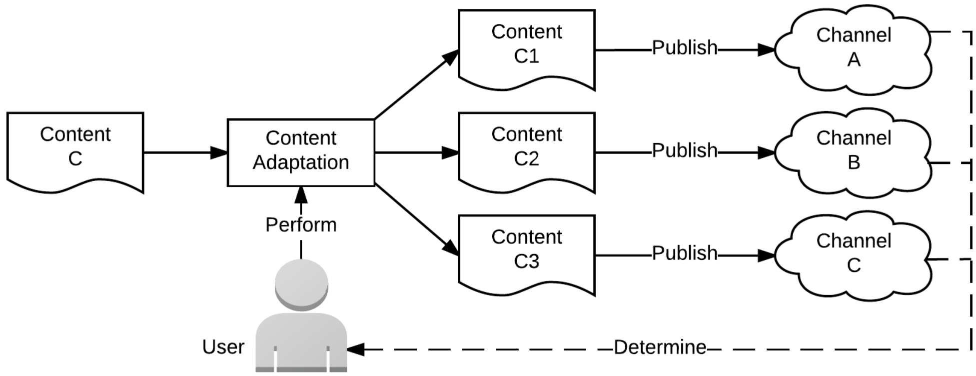

Figure 2: Content adaptation process in multi-channel online communication

Due to the diversity of knowledge possessed by an individual user in performing the adaptation, each adaptation might produce different output content. Publishing different adapted output contents might also attract different user engagements. By measuring the correlation between input and adapted output content, we will be able to measure the impact of the adaptation to the obtained user engagement as:

i. an adaptation transforms an input content and produces one or more output contents,
ii. each transformation will affect the quality of the produced content,
iii. publishing contents with different quality will attract different engagements,
iv. an adaptation will ultimately affect the user engagements.

Capability to understand the impact of the adaptation process to user engagements is an important aspect in multi-channel online marketing, mainly to:

i. have a better understanding of how to perform an adaptation such that a high quality output is able to produced for different channels,
ii. be able to automate the adaptation process by understanding how users performed the adaptation such that can be replicated.

This understanding could benefit each organization, namely to increase their content engagement on social media by designing their content adaptation processes carefully and automatically.

2.1 Vector Space Model

As previously described, a different adaptation might produce different output content that could attract different engagement. A way to measure these differences is by representing the input and output content using the same measurable representation, such that their similarities (or differences) can be computed. In this work, we use the vector space model (Salton et al. 1975) to compute the similarities between contents through their corresponding terms and term weights. Each content will be represented by a *t*-dimensional vector $D_i = (d_{i1}, d_{i2}, \ldots, d_{it})$, where d_{ij} represents the weight of *j*-th term. Using this representation, an angle between corresponding vectors can be computed as similarity measure, where a smaller angle represents a higher similarity.

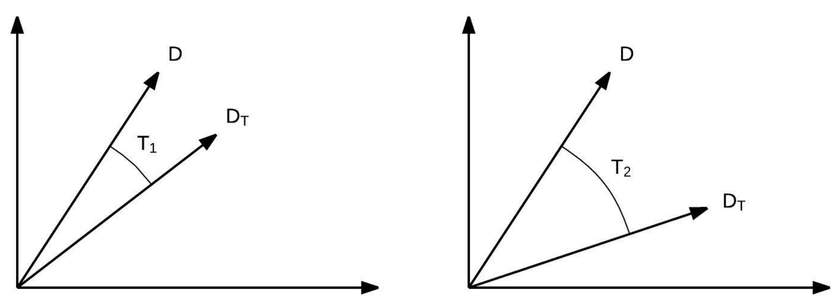

Figure 3: Vectors representation of input content D and output content D_T by using different adaptation T_1 and T_2

In the case of content adaptation, the measure is used to measure the similarity between the source and adapted content (e.g. a post or tweet) which is published on a social media channel. As shown in Figure 3, two different content adaptations T_1 and T_2 were applied to a content D and produced an output content D_T. Each adaptation might produce different similarities in degrees compared to the source as represented with different angles, as illustrated in Figure 3(a) and Figure 3(b) respectively.

We represent the adaptation impact to an input content based on the similarity degrees produced by this model. Adapting an input content to a specific attributes and or requirements will affect the similarity in degrees of the produced output.

2.2 Content Representation and Adaptation

We have identified three different content representations that could be affected by the content adaptation:

i. *textual*, which is related to how a content is presented textually,
ii. *hypermedia*, which is related to what kind of media are included in a content,
iii. *named-entity*, which is related to what kind of named entities are included in a content.

Textual As the most commonly used representation, a content is seen as a collection of terms (words or compound words) expressing specific meanings. An adaptation can be performed by altering members of the collection which is ultimately will also alter the conveyed meanings.

Hypermedia Hypermedia (Hardman et al. 1994) is an extension of Hypertext (Halasz and Schwartz 1994), where information is composed of hierarchy of data containing various “component” (e.g. text, graphics, audio, and video) interconnected by relational “links”. Content can be represented as a collection of these components and an adaptation can be performed by altering members of the collection.

Named-entity Named entities are phrases (a word or group of words) that contain the name of persons, organizations, locations, times and quantities (Sang 2002). The collections of these entities can be used to represent content such that an adaptation can be performed to a collection by altering its members. For example, the text “The first Winter Youth Olympics Games were held in Innsbruck, Austria in 2012” contains two named entities: “Innsbruck, Austria” as location and “2012” as date.

2.3 Content Adaptation and User Engagement

The impact of content adaptation can be measured by computing the similarity degree between the source and adapted content using previously explained representations. With assumption that a different adaptation might produce a different similarity degree, and publishing contents with different similarity degrees will acquire different types of user engagement, we can compute the correlation between content adaptation and user engagement. This relation between content adaptation T and user engagement E on content D can be formalized as follows:

i. An adaptation $t \in T$ on a content $d \in D$ produces a similarity degree θ,
ii. Publish the content d to a social media channel attracts an engagement $e \in E$,
iii. The most engaged contents $\{d \in D \mid e(d) \geq n\}$ are computed by using three content representations explained above, where n is a threshold constant.
iv. The similarity degree θ is then compared on different engagement metrics, channels and cases.

Since there are many user engagement measures available on each social media channel, we restrict our engagement measurement as follow: (i) the number of likes and comments on Facebook, (ii) the number of re-tweets and favorites on Twitter. Obviously, this measurement list can be extended and expanded to different available metrics on many social network sites. Our focus in this paper is more on the relation between content adaptations to user engagement which can be represented by any social media engagement measures.

3. Experiment and Result

To obtain the patterns of content adaptation impact on social media, we use the Facebook and Twitter channels of several tourist associations in Austria. This is simply because the touristic industry is a global industry where their targets are coming from around the globe, and social media channels are the major means to reach those audiences. In principle, the channels could be any online communication channels from any different industry. First, we explain how to obtain the required data, followed by how to compute the three similarities using various existing libraries, and finally we show and explain the obtained results.

3.1 Experiment

To be able to measure the transformation, we selected content posted on social media channels which were originated from a website, such that we can compare a post or tweet to its origin website as the source. The experiments were performed as follow:

1. *Collecting*. Posts and tweets from the selected channels were extracted and collected including their user engagement measures (Likes/Comments for Facebook's posts and Re-tweet and Favorite for Twitter's tweets). The collecting process was performed by utilizing the Facebook and Twitter APIs.
2. *Filtering*. The obtained posts/tweets then filtered, where only posts/tweets were originated from a website sources are selected. On Facebook they can be identified by the posting type *link* and *video*, where the link to the original sources is available in property *link*. On Twitter they can be identified through property *urls[]*
3. *Content Representation*. Each post/tweet then represented in textual, hypermedia, and named entity. A post consists of several properties, namely {*message, name, description, picture and link*}, while a tweet consists of {*text and urls[].url*}.
    a. In textual representation, posts/tweets are computed by using Bag-of-Words model[1].
    b. In hypermedia representation, each media in posts/tweets is detected using Mime Types Detector of Apache Tika[2]. 5 media types were detected based on the Internet media types specified in RFC 2046[3], namely *text, image, audio, video, and application*.
    c. In named entity representation, we use Stanford Named Entity Recognizer (NER)[4] to detect 7 named entities: *Time, Location, Organization, Person, Money, Percent, and Date*.
4. *The Most Engaged Content Selection*. For each representation, posts/tweets are mapped to their engagement measures. The most (or second, third, and so on) engaged contents are then selected.
5. *Similarity Computing*. The similarity degrees between the selected posts/tweets to their sources can be computed by applying the similar representation in step 3 to the sources.
6. *Mapping*. The obtained similarity degrees from the most engaged posts/tweets to their sources then visualized and compared.

Dataset In our experiment, we collected the adapted contents from two of the most popular social media platforms (Facebook and Twitter). We selected social media channels from three tourist associations (Tourismusverbande (TVb)) from Tyrol, Austria, namely TVB Innsbruck and its Holiday Villages[5], TVB Osttirol[6], and TVB Pillerseetal[7]. The experiment was performed at the end of December 2014, where 2946 posts from three Facebook channels and 4017 tweets from three Twitter channels were successfully collected. After filtering out the unnecessary posts/tweets (e.g. has no link to the original sources or the URL sources are inaccessible) and make sure the similarity degree for three representations can be computed successfully, at the end we have two datasets for Facebook posts (contains 643 posts) and Twitter tweets (contains 3778 tweets).

3.2 Result

By employing the three content representations explained above, we obtained three distinctive engagement mappings for each Facebook posts and Twitter tweets datasets.

[1] http://en.wikipedia.org/wiki/Bag-of-words_model
[2] http://tika.apache.org
[3] https://www.ietf.org/rfc/rfc2046.txt
[4] http://nlp.stanford.edu/software/CRF-NER.shtml
[5] http://www.innsbruck.info/
[6] http://www.osttirol.com/
[7] http://www.pillerseetal.at/

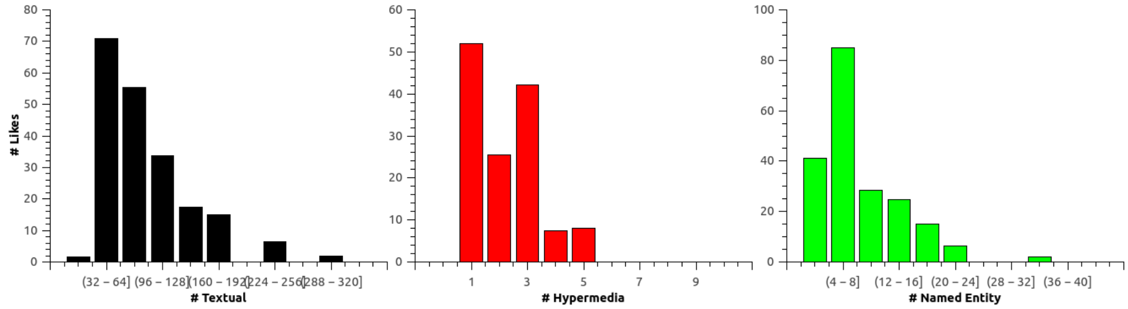

Figure 4: Comparison of Facebook's posts in three representations to the number of likes

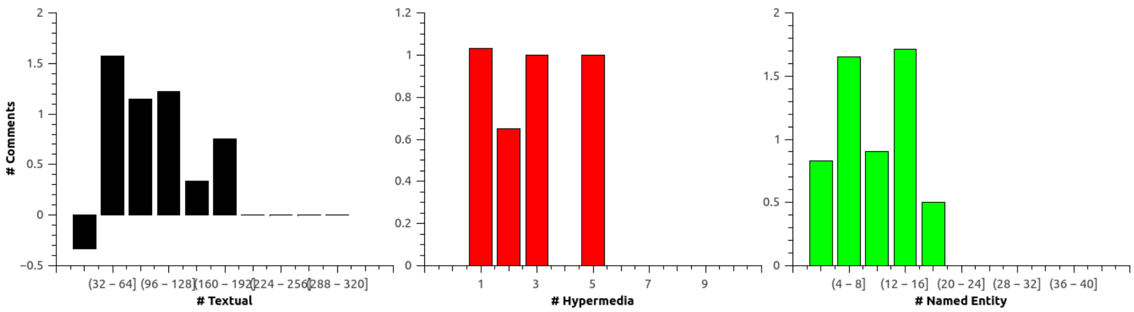

Figure 5: Comparison of Facebook's posts in three representations to the number of comments

Figure 4 and Figure 5 show a comparison between posts in textual, hypermedia and named entity representations to the number of likes and comments. For the likes engagement, the highest average number of likes (>70) was obtained within the range of (32-64], 1 (>52), (4-8] (>84) in textual, hypermedia and named entity representation respectively. For the comments engagement, the highest average number of comments (>1.6) was obtained within the range of (32-64], 1 (>1.03), (12-16] (>1.7) in textual, hypermedia, and named entity representation respectively. Both figures show similar patterns for textual and hypermedia representations. Posts with high engagement are spread within a narrower area in hypermedia representation (1-3) compared to the other representations. Posts with more named entities are not necessarily obtaining more engagement.

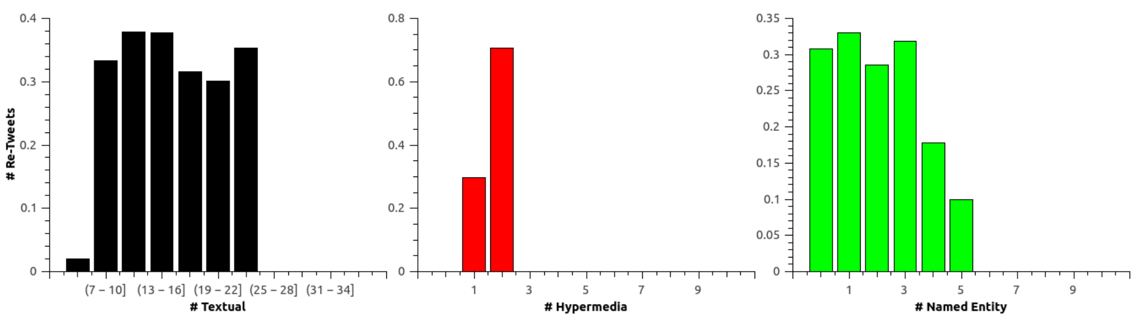

Figure 6: Comparison of Twitter's tweets in three representations to the number of re-tweets

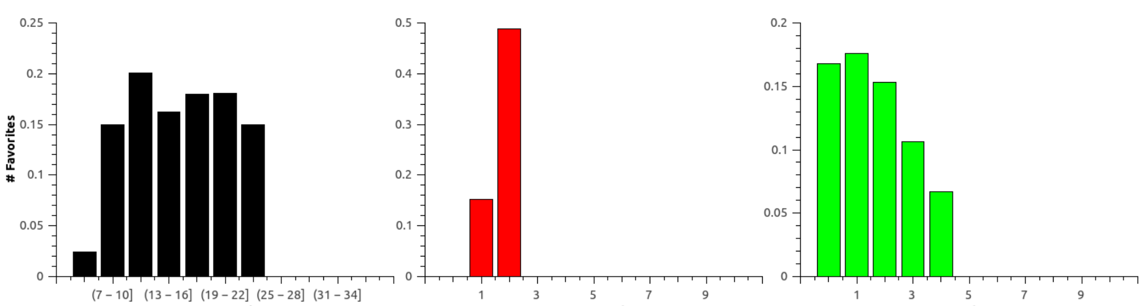

Figure 7: Comparison of Twitter's tweets in three representations to the number of favorites

Figure 6 and Figure 7 compare tweets in textual, hypermedia and named entity representations to the number of re-tweets and favorites. For re-tweets engagement, the highest average number of re-tweets (>0.38) was obtained within the range of (10-13], 2 (>0.7), 1 (>0.33) in textual, hypermedia, and named entity representation respectively. For favorites engagement, the highest average number of favorites (>0.2) was obtained within the range of (10-13], 2 (>0.49), 1 (>0.18) in textual, hypermedia, and named entity representation respectively. Both figures show similar patterns for each representation. Tweets with high engagement are spread within a narrow area (1-2) in hypermedia representation, wider area (0-5) in named entity representation, and widest area (0-25) in textual representation.

Given those ranges, we were able to compose rules to select the most engaged posts/tweets based on the three content representations as shown at Table 1. Of course, the rules can be composed from the second, third, and so on most engaged contents. We use the most engaged ones solely for simplicity.

Table 1: Rules extracted from the most engaged posts and tweets

Rule	Textual	Hypermedia	Named Entity
Facebook	(32-64]	1	(4-8] or (12-16]
Twitter	(10-13]	2	1

Based on these rules, we selected the most engaged posts/tweets from Facebook (356 posts) and Twitter (1582 tweets) datasets. The similarity degree between each selected post/tweet to its source then computed and compared.

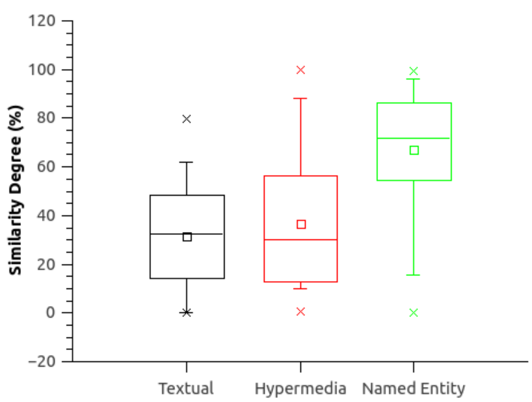

Figure 8: Similarity degree distribution of the most engaged Facebook's posts

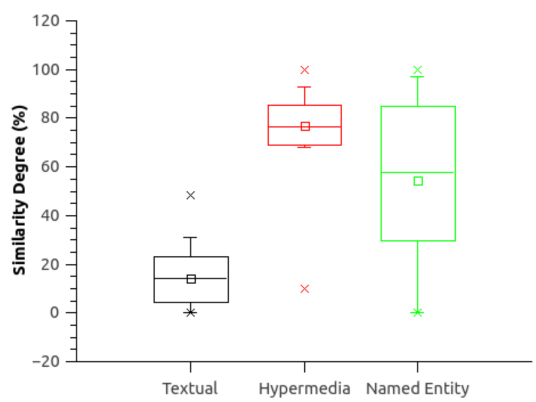

Figure 9: Similarity degree distribution of the most engaged Twitter's tweets

Figure 8 and Figure 9 show the similarity degree comparison of the selected posts and tweets to their sources in three content representations. The interquartile ranges for posts are 15-48% (the median is about 33%) in textual representation, 13-57% (the median is about 30%) in hypermedia representation, and 55-87% (the median is about 72%) in named entity representation. For tweets, the interquartile ranges are 5-23% (the median is about 14%) in textual representation, 70-85% (the median is about 77%) in hypermedia representation, and 30-85% (the median is about 58%) in named entity representation.

Comparing both plots, there are a few strong indications that:

i. In textual representation, the similarity degrees between posts to their sources are largely less than 50% and for tweets even lower (below 30%). Obviously, this case is reasonable since a tweet is limited to a maximum of 140 characters.
ii. In hypermedia representation, most of the tweets are highly similar to their sources, compared to the posts. A tweet is typically accompanied by more than one link.
iii. In named entity representation, both datasets show a similar distribution, where most of the posts/tweets have a high similarity to their original sources (above 50%).

3.3 Validation

To validate and proving the usefulness of the proposed evaluation procedures, we applied the similar procedures to a different dataset. We collected Facebook's posts from ten hotels located in Austria and computed with three representations explained above and compared to the number of likes and comments. The results are shown in Figure 10 and Figure 11.

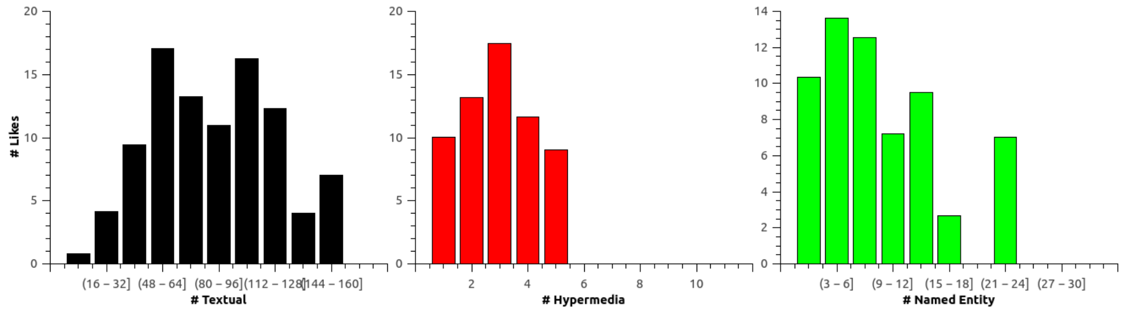

Figure 10: Comparison of Facebook's posts from ten hotels to the number of likes

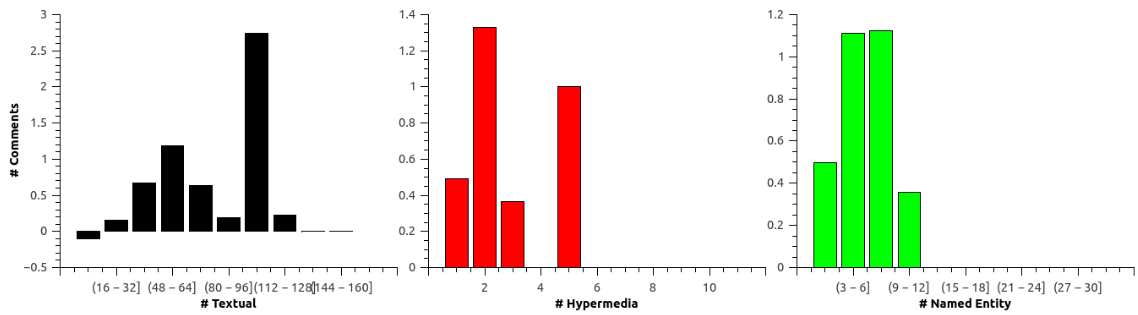

Figure 11: Comparison of Facebook's posts from ten hotels to the number of comments

Based on the comparison results, a rule was composed to select the most engaged posts. The highest number of likes is within the range of (48-64] and the highest number of comments is within the range of (96-112] in textual representation. In hypermedia representation, the highest number of likes and comments are 3 and 2 respectively. In named entity representation, the highest number of likes is within the range of (3-6] and the highest number of comments is within the range of (6-9]. By using this rule, from 436 posts in this dataset, 262 of them were selected.

Figure 12 shows the similarity degree distribution of the selected posts. The interquartile ranges are 13-49% (the median is about 31%) in textual representation, 28-76% (the median is about 46%) in hypermedia representation, and 54-91% (the median is 76%) in named entity representation.

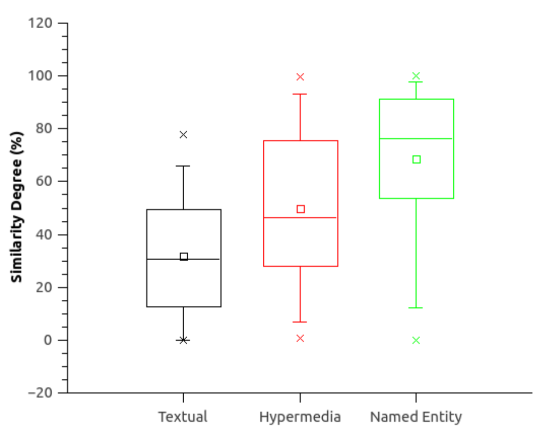

Figure 12: Similarity degree distribution of the most engaged Facebook's posts from hotel dataset

Compared to the first Facebook dataset, the similarity degree distributions in textual and named entity representations are relatively similar. The difference is in hypermedia representation, where median in the hotel dataset is 46%, higher than 30% from the first dataset. This is a strong indication that the hotels are more likely incorporated with more links (images, videos) in their posts.

3.4 Discussion

Our main intention is to be able to identify important relations between content adaptations (which are performed by humans) to the user engagement on social media. We constructed the relation by comparing the similarity of posts/tweets and their original sources to the obtained engagement. First, we selected the most engaged posts/tweets according to three different content representations. Then, the similarity degrees from the selected posts/tweets to their original sources are computed and compared.

First, from the two first datasets, even though the number of tweets is five times more than the number of posts, the average user engagement to the posts is very high: the average number of likes is in range 52-84 and the number of comments is in range 1.03-1.7, while the number of re-tweets is in range 0.33-0.7 and the number of favorites is range 0.18-0.49. It seems that the audiences on the Facebook channel are more active.

Second, from two Facebook datasets, we obtained similar patterns in textual and named entity representations but differ in hypermedia representation. This fact validates the usefulness of using various types of content representations, where datasets can be compared in numerous ways.

4. Related Work

There are at least two topics that relate to our work. First, the content adaptation for personalization access to various information sources, for example to provide users with a better experience and higher level of satisfaction when accessing information space, especially from social media (Arachchi and Dogan 2013). Adaptation is widely used for the adaptive hypermedia (Brusilovsky 1996, Brusilovsky 2001) which later became the adaptive web (Brusilovsky and Maybury 2002) which tends to fulfill the particular needs of individual users or a group of users in an optimized access to distributed information on a hypermedia system or the web. Adaptive semantic web (Dolog et al. 2003) tries to bring it a step further by employing standardized description formats for metadata to allow us to reason over facts described in the formats. In these cases, the content adaptation is performed to transform a content to fit them into user needs.

The second topic is related to methods and approaches to improving the engagement on social media, such as in (Suh et al. 2010, Hong et al. 2011, Gummerus et al. 2012, Pletikosa and Michahelles 2013, Bitter et al. 2014). In contrast to those works, our work is trying to understand the impact of content adaptation to the user engagement, such that we can improve the engagement by carefully adjusting the content adaptation.

5. Conclusion and Future Work

In this work, our main goal is to have a better understanding of how a content adaptation should be performed in a way suitable for publication on a particular social media channel and capable of reaching a wider audience, thus increasing user engagement.

We evaluated the adaptation impact to the user engagement by introducing and utilizing three different content representations (namely textual, hypermedia and named entity) to: (i) select the most engaged contents, (ii) compute the similarity degrees between original and adapted content. By comparing the obtained similarity degrees on different engagement metrics, channels, and cases, we acquired new knowledge on how the content adaptation can be adjusted to obtain higher user engagements. We tested our evaluation procedures on the social media channels of Facebook and Twitter from several tourism industries. We were able to identify the similarities and differences on social media posting strategies. In the future, we would like to extend our evaluation to other social media channels from various platforms, as well as from social media channels from different areas.

Finally, we are planning to incorporate the obtained engagement patterns into publication rules (Akbar et al. 2014a) and implement them into the multi-channel online communication platform (Akbar et al.

2014b) to determine a fine tuning content adaptation and to enable an automatic adaptation for high quality content dissemination on social media and online communication in general.

Acknowledgements
This work has been partially funded by project TourPack[8], supported by the Austrian Research Promotion Agency (FFG) within the program "Future ICT". We would like to thank all the members of the Online Communication working group[9] for their valuable feedback and suggestions.

References

Akbar, Z., Garcia, J.M., Toma, I. and Fensel, D. (2014a) “On using semantically-aware rules for efficient online communication”, In Bikakis, A., Fodor, P., Roman, D., eds.: Rules on the Web. From Theory to Applications. Volume 8620 of LNCS. Springer, pp. 37-51

Akbar, Z., Garcia, J.M., Toma, I. and Fensel, D. (2014b) “A semantic-based platform for efficient online communication”, In: Proceedings of the 13th International Semantic Web Conference (ISWC) 2014 Posters & Demonstrations Track, pp. 337-340

Arachchi, H. and Dogan, S. (2013) “Context-aware content adaptation for personalised social media access”, In Ramzan, N., van Zwol, R., Lee, J.S., Clver, K., Hua, X.S., eds.: Social Media Retrieval. Computer Communications and Networks. Springer London, pp. 305-338

Bitter, S., Grabner-Krauter, S., and Breitenecker, R.J. (2014) "Customer engagement behaviour in online social networks - the Facebook perspective", International Journal of Networking and Virtual Organization, Vol. 14, No. 1-2, pp. 197-220

Brusilovsky, P. (1996) “Methods and techniques of adaptive hypermedia”, User Modeling and User-Adapted Interaction, Vol. 6, No. 2-3, pp. 87-129

Brusilovsky, P. (2001) “Adaptive hypermedia”, User Modeling and User-Adapted Interaction, Vol 11, No. 1-2, pp. 87-110

Brusilovsky, P. and Maybury, M.T. (2002) “From adaptive hypermedia to the adaptive web”, Communication ACM, Vol. 45, No. 5, pp. 30-33

Dolog, P., Henze, N., Nejdl, W. and Sintek, M. (2003) “Towards the adaptive semantic web”, In Bry, F., Henze, N., Mauszyski, J., eds.: Principles and Practice of Semantic Web Reasoning. Volume 2901 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 51-68

Gummerus, J., Liljander V., Weman, E. and Pihlstrm, M. (2012) "Customer engagement in a Facebook brand community", Management Research Review, Vol. 35, No. 9, pp. 857-877

Halasz, F. and Schwartz, M. “The dexter hypertext reference model”, Communication ACM, Vol. 37 No. 2, pp. 30-39

Hardman, L., Bulterman, D.C.A. and van Rossum, G. (1994) “The amsterdam hypermedia model: Adding time and context to the dexter model”, Communication ACM, Vol. 37 No. 2, pp. 50-62

Hong, L., Dan, O. and Davison, B.D. (2011) "Predicting popular messages in Twitter". In Proceedings of the 20th international conference companion on World wide web (WWW '11). ACM, New York, NY, USA, pp. 57-58

Pletikosa Cvijikj, I. and Michahelles, F. (2013) “Online engagement factors on facebook brand pages”, Social Network Analysis and Mining, Vol. 3, No. 4, pp. 843-861

Salton, G., Wong, A. and Yang, C.S. (1975) “A vector space model for automatic indexing”, Communication ACM, Vol. 18, No. 11, pp. 613-620

Sang, E.F.T.K. (2002) “Introduction to the conll-2002 shared task: Language-independent named entity recognition”, In: Proceedings of CoNLL-2002, Taipei, Taiwan, pp. 155-158

Suh, B., Hong, L., Pirolli, P. and Chi, Ed H. (2010) "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," In Proceeding of the IEEE Second International Conference on Social Computing (SocialCom), pp. 177-184

[8] http://tourpack.sti2.at
[9] http://oc.sti2.at