Research Article

International Journal of Advanced Computer Research, Vol 7(31)
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2017.731001

Online collaborative video annotation framework using GoodRelations ontology for E-commerce

Triet H. M. LE and Hai T. Duong[*]
School of Computer Science and Engineering, International University-Vietnam National University, Ho Chi Minh City, Vietnam, Quarter 6, Linh Trung Ward, Thu Duc District, HCMC, Vietnam

Received: 01-March-2017; Revised: 27-May-2017; Accepted: 31-May-2017

Abstract

In recent years, E-commerce has become one of the fastest growing industry and constantly provided better services for the customers. Such incredible advance was partially thanks to the mature development of the semantic web technologies with better content support for the search engines and recommender systems. Moreover, with rich embedded information, hypervideo has turned out to be a promising way of online shopping. In order to facilitate its development, GoodRelations ontology has been developed to make the video more comprehensive for both the users and machines. However, due to the vastness and diversity of the internet, the existing methodologies could not reach a consensus on the information annotated by the participants. Therefore, in this study, we propose a semantic video annotation framework using consensus quality of the collaborative process and GoodRelations ontology as the domain knowledge. This approach has been demonstrated to enhance the quality of the annotated information, but still maintain the flexibility for the vendors to sell their products interactively on the videos. As a result, this study provides a robust and reliable video annotation platform to be used in the E-commerce industry.

Keywords

E-commerce, GoodRelations ontology, Video annotation, Collaborative framework, Consensus quality.

1.Introduction

Since the introduction of the World Wide Web (WWW) in 1989, it has revolutionized every aspect of the people's life dramatically. With the aid of hypertext markup language (HTML), the users could view the content of the websites with ease. However, shortly after the emergence of HTML, there was a serious lack of semantic within the web content. Hence, there have been many efforts to integrate the meaning into HTML tags such as Microformats, RDFa, Microdata and JSON-LD, etc. The most important component of these solutions is ontology. With the use to provide domain vocabularies, ontology has also been applied into E-commerce to support the identification of the products as well as the transactions performed by the customers [1]. One of the most popular ontologies used in E-commerce was developed by Martin Hepp namely GoodRelations [2] and later adopted by schema.org (previously known as www.data-vocabulary.org) in 2011. The reputation of this ontology has been established to its flexibility to publish and share the details of both the products and the services with minimal requirements. It is also very friendly to the search engines and web crawlers. Besides, the semantic web has been particularly useful to build intelligent recommender systems using the hidden patterns derived from the content of the E-commerce websites. Therefore, thousands of E-commerce websites, including some renowned ones such as Amazon, Bestbuy, eBay, etc. exploiting the power of GoodRelations as their domain knowledge to support their business.

Recently, Duong et al. conducted a novel study on the use of GoodRelations ontology in video annotation to serve as a new way of online shopping [3]. Instead of accessing an E-commerce site to search for a particular product like the traditional way, the clients and vendors interact directly with each other on the videos. More specifically, the vendor annotates the products appearing in the video at several particular frames along with the related information. On the other side, the customers can view and make purchases similarly to using the

[*] Author for correspondence

product page of the conventional E-commerce sites. Thanks to the integration of GoodRelations ontology into the annotations, the annotators can follow some predefined semantic definition to provide relevant information for the clients to do better shopping.

However, the author did not address the vagueness and uncertainty of the semantic web. More clearly, the content making use of the semantic web technology is not guaranteed to follow a rule strictly due to the diverse nature of the human's expressions. As a result, in this paper, we propose a new way to devise a framework to apply the collaborative process based on consensus quality to overcome the existing challenges of the video annotation in E-commerce. The remaining structure of our paper is organized as follows. Section 2 summarizes the current studies on semantic video annotation techniques as well as the collaborative algorithms. Section 3 presents our proposed collaborative video annotation framework based on consensus quality. Section 4 demonstrates some of our experimental results and section 5 draw conclusions on our work.

2.Related works

The concept of information integration on videos has been proposed since the 1980s [4]. The research on hypervideo was then extended to the term "clickable video" in [5]. Some available products have been introduced to turn the idea into reality such as *VideoClix*[1], *#HAL*[2] (formerly known as *ClikThrough*), and *LinkTo.Tv*[3]. Both *VideoClix* and *#HAL* offer the user the ability to create multiple object annotations within the uploaded video. While watching the video, the user can interactively view the annotated object of interest by hovering the mouse over them, and a brief piece of related information is then displayed to the user. Similarly, *LinkTo.Tv* allows the user to select the keyframes to create the hot spots as object annotations with additional information. After the annotation process is complete, the user can share their annotated videos on their channels or other social networks. One popular approach to semantic video annotation was through the advances in computer vision and machine learning algorithms [6-8]. The main idea is to extract the essential features in the frames to make inferences about high-level concepts existing in the video. One such pioneer product was *wireWAX*[4]. Firstly, it provides both automatic face detection and semi-automatic object identification based on user's previous annotations. Subsequently, the system suggests the boundaries of each object to support hypervideo. Apart from commercial releases, there have also been several well-known video annotation tools and datasets such as Anvil [9] and Label Me [10]. In the former scheme, the annotation is in the form of a variable-size rectangle with either linear interpolation or dynamic tracking. The tracking was implemented using the support vector machine based on histogram of oriented gradients (HOG) and color features. In the latter study, the authors proposed an interesting method namely homography-preserving shape interpolation to minimize the error during the object movement caused by 3D to 2D mapping. However, except for some extreme cases when the trajectory of the object is curve-like as in *Figure 1*, the linear interpolation is sufficient to approximate the movement of objects without bearing too much burden on the pixel manipulation. As a result, in this paper, we adopted the simple linear approach to keep track of the movement of the object annotation.

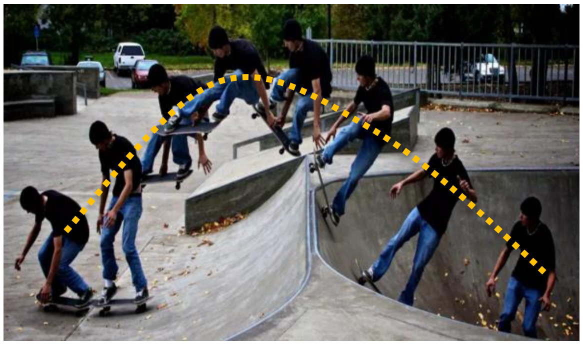


Figure 1 Curve-like movement of the skateboarder, which is difficult to use linear interpolation to annotate correctly

However, the common challenge of existing video annotation tools is that they do not support the user to work in an in-depth collaborative environment. Although they have provided some mechanisms to let the user share their annotations within their network, there is no concrete way to resolve the conflicts between different versions of a shared annotated object. In fact, there have been some efforts to build the system architecture to facilitate the collaboration among users [11-13].

[1] VideoClix: Clickable Video, Interactive video advertising https://web.archive.org/web/20140704033625/http://www.videoclix.tv/. Accessed 15 January 2017.

[2] #HAL: Video intelligence http://www.hashtaghal.com/. Accessed 15 January 2017.

[3] LinkTo.Tv - Interactive video made simple https://www.linkto.tv/. Accessed 15 January 2017.

[4] WIREWAX - No.1 Interactive Video Technology http://www.wirewax.com/. Accessed 15 January 2017.

In most cases, a concept or domain ontology was designed to match the retrieved features with existing terms to determine the description of the object. Such approaches generated a high motivation to develop a robust multimedia ontology to connect between low-level features with high-level knowledge [14-16]. Among them, GoodRelations ontology [2] is highly appropriate to support E-commerce video annotation environment which will be discussed in Section 3.1.

Besides the domain knowledge, there is also a need for an underlying framework to compromise various forms of a single annotation. There are two common methodologies to fulfill this goal: nominal group technique (NGT) [17] and Delphi method [18]. The former approach requires an in-person meeting between the experts to reach the final common goals taking into consideration the contribution of each participant. On the contrary, the latter method supports anonymous collaboration to achieve the consensus among the users. Although the NGT usually takes less time than Delphi method, it is difficult to meet directly; especially, in the case people only connect with each other via a social network. In an E-commerce environment, direct communication between users is sometimes impossible. As a result, the Delphi method can be utilized to build the framework to allow broader collaboration among the vendors.

3.Online collaborative video annotation framework using GoodRelations ontology based on consensus quality

Our main philosophy of our framework is to create a common E-commerce platform for both the clients and suppliers. In other words, the vendors draw the annotations for the products appearing within the video that they want to sell. On the other side, the customer can view the information of those annotations while watching the video. In this way, there is a direct channel to connect the clients and the suppliers to make the most beneficial transactions for both sides. The overall framework of our proposed solution is illustrated in *Figure 2*. As mentioned in section 2, the GoodRelations ontology [2] provided the common vocabularies for labels during the annotation process. In this study, the labels represented the attributes of the product annotated by the users. The details will be discussed in section 3.1. In the next step, the user created the annotation in term of the polygon using our tool with the pre-defined labels, of which details will be described in section 3.2. Subsequently, the user could share their annotations via their social network of choice so that other people could join them to contribute to the same annotated object/product. Using the consensus quality approach introduced in section 3.3, our system was able to recommend the users with the collaborative version from earlier users' annotated information when they started a new annotation process. The whole process was repeated until all of the objects of interest had been annotated.

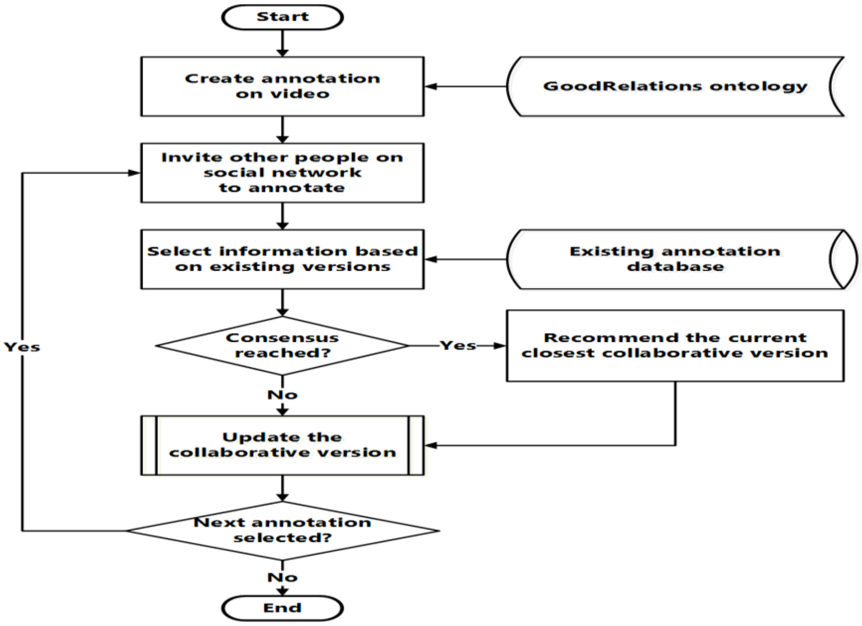

Start
Create annotation on video
GoodRelations ontology
Invite other people on social network to annotate
Select information based on existing versions
Existing annotation database
Consensus reached?
Yes
Recommend the current closest collaborative version
No
Update the collaborative version
Yes
Next annotation selected?
No
End

Figure 2 Overall flowchart of our proposed framework

3.1Overview about the GoodRelations ontology and schema.org vocabulary

Traditionally, the relational database is commonly used to store data with simple relationships. On the other hand, ontology can represent a higher level of knowledge and model with much more complex dependency among the entities for any system concerning about the semantic meaning [19]. As a result, we can make inferences based on the rules defined in the ontology to unveil hidden knowledge, which cannot be accomplished using only the schemas of the relational database.

In the case of E-commerce, GoodRelations [2] has long been considered a robust ontology to provide the vocabulary for the semantic web. According to its class diagram[5], there are 27 classes with many properties and hundreds of relationships among the entities. As the content on the internet grew at an unprecedented pace, there was an inevitable need for more terms in the ontology to express the meaning of new concepts. With GoodRelations ontology at its core, schema.org vocabulary extended the number of classes to 315 with a much greater number of properties and relationships[6]. With such variety of structure, schema.org can describe a broader range of product/service information, transactions, movies, images, etc. Moreover, it has been designed to work with several semantic web languages such as Microdata, RDFa and Turtle, etc. Thanks to this support, the effectiveness of the search engine optimization techniques has increased considerably. In addition, the quality of the answers to the user's queries has been constantly improved to meet everyone's needs to the fullest. With the acknowledgement of such advantages of both GoodRelations ontology and schema.org, in our study, we adopted some of its concepts to describe the attributes of the annotated products/objects. In this way, we could both allow the search engine to recognize the meaning of the user's annotations and provide a common ground to achieve consensus in the collaboration.

3.2Online video annotation tool

For this module, we mainly used Canvas element of HTML5 to display the annotation as a separate layer upon the running video. With the aid of JavaScript, our system could handle the movement of the annotations (cf. *Figure 3*). In addition, the user could display/hide a specific or multiple annotations in the current video.

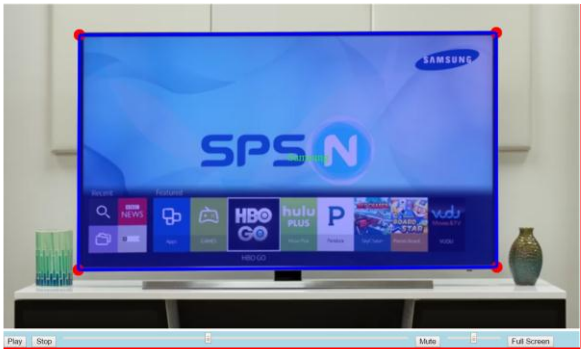


Figure 3 A polygon in form of a rectangle created to annotate the TV as an upper layer on the video using our tool

The object was polygon made up of three or more points in the 2D coordinate system, offering more flexibility for the user than the method introduced in [9]. Moreover, our tool provided the user with the capability to create the annotation faster by just using double click. In such case, the double click created a single dot on the canvas to represent the object (cf. *Figure 4*), which required less effort than drawing a whole polygon-like boundary.

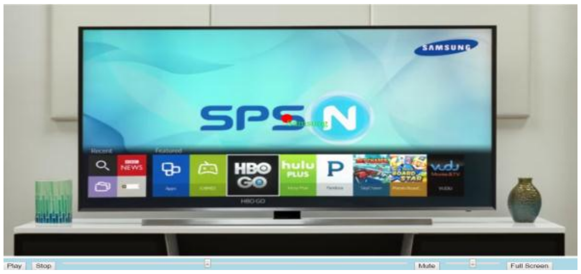


Figure 4 A simple red dot at the center of the TV appearing in figure 3 to replace the previous rectangular annotation, reducing the effort for the annotators

This approach turned out to be more convenient than the one proposed in LabelMe tool [10] during the annotation process. The user sometimes did not pay too much attention to the precise boundaries of the object, but instead they just wanted to show its existence within the video quickly. With this method, we did not take into consideration complex object tracking in the video, which could save a considerable amount of internet bandwidth and computational resources. More specifically, the

[5] UML class diagram of the GoodRelations ontology http://www.heppnetz.de/ontologies/goodrelations/goodrelations-UML.png. Accessed 15 January 2017.

[6] GoodRelations and schema.org http://wiki.goodrelations-vocabulary.org/GoodRelations_and_schema.org. Accessed 15 January 2017.

annotation moved in a linear fashion according to the starting/ending frames and time specifications of the user. The movement of the whole polygon was possible thanks to the change in the location of each point. In our proposed approach, each 2D point was transformed using the linear interpolation in (1).

$$x' = x_0 + \frac{(f' - f_0)}{t' - t_0} \qquad \text{for } 0 \le t_0 \le t' \qquad (1)$$
$$y' = y_0 + \frac{(f' - f_0)}{t' - t_0}$$

Where x_0, y_0, f_0, t_0 were the initial x-coordinate, y-coordinate, video frame and time, respectively; whereas, x', y', f', t' were the x-coordinate, y-coordinate, video frame and time when there existed a sudden change in the movement direction of the selected object. Equation (1) also specified that the user seldom needed to modify the positions of the object annotation. Specifically, our framework automatically determined the position of the polygon between two key frames, speeding up the annotation process in overall. However, a challenge exists during the annotation process on how to determine whether the position of the current cursor of the user's mouse fell within the region of any annotated object. If the user was selecting the annotated object, the boundary of such object would be displayed on the canvas for him/her to view, add or modify the object information. In order to achieve such purpose, an algorithm was developed based on the even-odd rule algorithm (also known as the crossing number algorithm) [20]. In this work, the even-odd rule was also extended to the case of non-polygon (e.g., single point or straight line) where the numbers of points were less than three. The result of the whole annotation process would then be saved to an XML file. The XML schema definition of the XML-based annotation file is described in Figure A1 in appendices. Each annotated object was assigned a unique id for its identification along with some important information about its existence in a given video (e.g., the time of appearance/disappearance, the keyframes of movement and different set of attributes input from multiple users). With this file, when the user loaded the same video the next time, all of the annotated objects would appear the same as the previous time using the parsing module implemented inside our system. In more detailed, in order to match the XML elements with the objects in the video, we followed three steps.

First, we parsed the XML file using document object model to extract the elements. Then, we saved these attributes into the corresponding object (i.e., appearing/disappearing frame, keyframes, hidden frames, static frames). Finally, to display these annotations, we continuously redraw the canvas after a very small pre-defined interval (e.g., 30ms) and simultaneously check the position and status of each object (e.g., moving, hidden). By this way, we could determine to display which object at which location on the canvas upon the running video.

3.3Collaborative framework for video annotation using consensus quality

Unlike some existing works in collaboration such as WordNet [21, 22], where all of the concepts need to be compromised to reach a single final version to share with the community. In an E-commerce environment, we need to balance between the common knowledge and the freedom for vendors to advertise their products. Hence, we only proposed the collaborative process for the shared characteristics of a single product such as its name, model, description and brand as described in *Figure 5*.

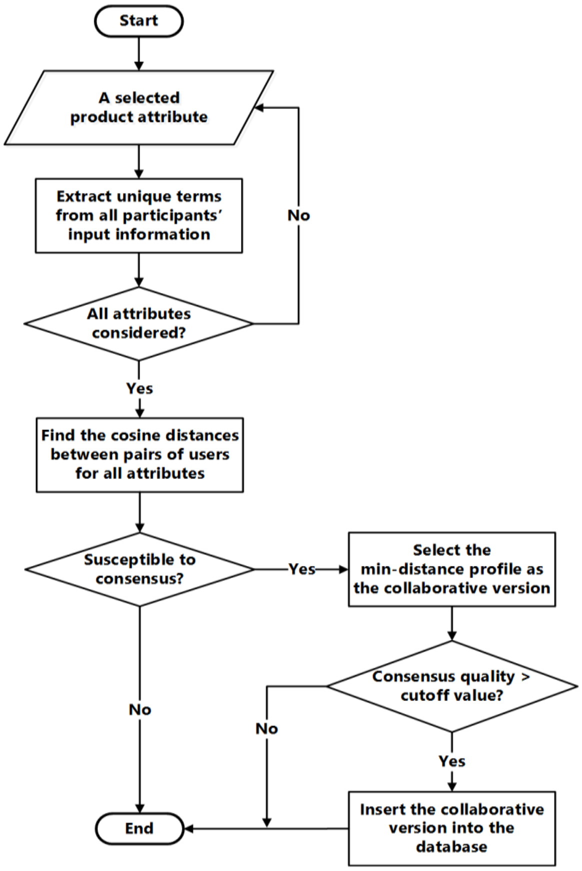

Start
A selected product attribute
Extract unique terms from all participants' input information
No
All attributes considered?
Yes
Find the cosine distances between pairs of users for all attributes
Susceptible to consensus?
Yes
Select the min-distance profile as the collaborative version
Consensus quality > cutoff value?
No
No
Yes
Insert the collaborative version into the database
End

Figure 5 Our proposed collaborative framework based on consensus quality for video annotation in E-commerce

At the beginning of the collaborative process, when an attribute of the product was selected, the system needed to figure out the unique terms from all the available versions of the users. The WordNet building process in [22] considered the whole phrase input by the user as a vector element used for the calculation of the cosine distance. However, it was nearly impossible to apply the same scheme in E-commerce environment due to the complexity and diversity of the words provided by the participants. Therefore, in this study, we applied binary weighting bag-of-words model to reduce the possibility of large distance due to different word order or small discrepancy in the way of language expression. It is noted that we focused more on extracting common information from multiple inputs than clustering them. Hence, we were more concerned about the appearance of a word than its frequency or tf-idf weight. In this sub-process, the attribute data were firstly combined from all of the user's versions into one piece of text. After that, seven steps were carried out by our system to clean the data including (1) replace all newline separators with spaces; (2) split all of the words separated by one or more spaces; (3) remove the punctuations (e.g., ",", ";", ".", etc.) to get the bare words; (4) decapitalize all of the characters to remove the case-sensitivity of the words; (5) remove all duplicated word(s) in the list; (6) eliminate all of the stop words (e.g., like, I, you, the, etc.); and (7) store all of the processed words in a list for later steps. The list of processed words was regarded as the anchor to calculate the cosine distance between any two users. The current process would repeat until the cosine distances all of the predefined attributes had been determined. Subsequently, the vector of each attribute was combined into a single vector for each user, in which the order of each attribute was still maintained. By this way, we could calculate the distance between two participants and determined the characteristics of the collaborative process based on the O1-consensus of Nguyen's study [23]. In particular, the criteria for *susceptible to consensus* as well as the quality of the collaborative process were established. Firstly, we assumed *n* to be the number of participants in the current collaborative process. Each participant in the list contributed their knowledge *x* to a predefined profile $\mathbf{X}_o$.

$$\mathbf{X}_o = \{ x_1, x_2, ..., x_n \}$$

where x_i is a combination of all input values of annotated object *o* based on the vocabulary provided by GoodRelations ontology [2] and schema.org of the i^{th} participant; whereas, $\mathbf{X}_o$ is the vector containing the knowledge of all participants in the collaborative process for the object *o* in the video.

Secondly, to resolve the conflicts during the annotation process among the participants, we needed to calculate their cosine distances, which represented the degree of discrepancy.

$$d(x_i, x_j) = 1 - cos(\theta) = 1 - \frac{\mathbf{A}.\mathbf{B}}{\|\mathbf{A}\|\,\|\mathbf{B}\|}$$

$$= 1 - \frac{\sum_{k=1}^{n} a_k b_k}{\sqrt{\sum_{k=1}^{n} a_k^2}\,\sqrt{\sum_{k=1}^{n} b_k^2}} \qquad (2)$$

where $d(x_i, x_j)$ is the distance between the annotated information x_i and x_j ($x_i, x_j \in X$) of i^{th} and j^{th} participants; **A**, **B** are the feature vectors and $\|\mathbf{A}\|$, $\|\mathbf{B}\|$ are the magnitudes of these vectors of x_i and x_j; a_k, b_k are the components of two vectors **A** and **B**, respectively; $cos(\theta)$ is the similarity between vector **A** and **B**. Moreover, we needed to calculate the total average distance of all distances in profile $\mathbf{X}_o$ ($d_{t\_mean}(X)$), the average distance of all distances between each annotated information *x* and profile $\mathbf{X}_o$ ($d_x(X)$), and the consensus value of the whole profile $\mathbf{X}_o$ (d_{min}) defined in (3-5), respectively.

$$d_{t\_mean}(X) = \frac{\sum_{x,y \in X} (d(x,y))}{n(n+1)} \qquad (3)$$

$$d_x(X) = \frac{\sum_{y \in X} (d(x,y))}{n} \qquad (4)$$

$$d_{min} = min_{u \in U}\, d_x(X) \qquad (5)$$

where *u* is a particular participant in a set *U* containing all of the participants in the profile $\mathbf{X}_o$.

Finally, after all of the necessary parameters were calculated, we needed to define the criterion for the *susceptible to consensus* to happen. According to [23], such condition occurred when the condition in (6) is met. At that time, the knowledge of profile would be dense enough to generate a sufficiently good compromise among all participants' annotations.

$$d_{t\_mean}(X) \geq d_{min}(X) \qquad (6)$$

Finally, another metric was required to determine the consensus quality denoted as $\hat{d}(x,X)$ of the current collaborative efforts among all of the participants as in (7). Using this measure, we could compare the quality of different profiles.

$$\hat{d}(x,X) = 1 - \frac{d(x,X)}{n} \quad (7)$$

Where $d(x,X)$ are the distances of the consensus to all other elements in the current profile $\mathbf{X}_o$.

The bottom line of this proposed framework was to reduce the effort of the annotation process and reach the agreement for the common attributes of each annotated object. As a result, in case condition for consensus susceptibility was reached, we chose the user with the minimum average cosine distance to all other participants in the current profile as the collaborative version. We could then recommend such knowledge to the new participants annotating the same object to speed up the input process for the common attributes. More importantly, the use of the consensus definitely enhances the quality of the current profile and helps achieve the common knowledge much faster [23]. We utilized this idea to make our system more robust based on the consensus quality. Furthermore, if the quality of a profile was larger than a cut-off value (e.g., 0.7), such collaborative version would be stored in the database to recommend for new annotations in the future.

4.Experiments

4.1Dataset

For this experiment, we used two sets of labels for each annotated object. The first one contained four pieces of common information for the product: (1) name, (2) model, (3) description and (4) brand. We also provided four more attributes for the vendors to distinguish themselves from the others including (1) name of the store, (2) link of the vendor's website, (3) product price and (4) rating. These vocabularies were taken from the concepts of schema.org[7] (i.e., previously GoodRelations ontology). The experiment was conducted on a particular object (e.g., a television) with 24 rounds of contribution. Specifically, many people were invited to annotate a same television on the video.

From the links provided by the participants, it turned out that the information of the product mainly came from popular E-commerce websites such as Amazon, eBay, Best Buy, Walmart, etc.

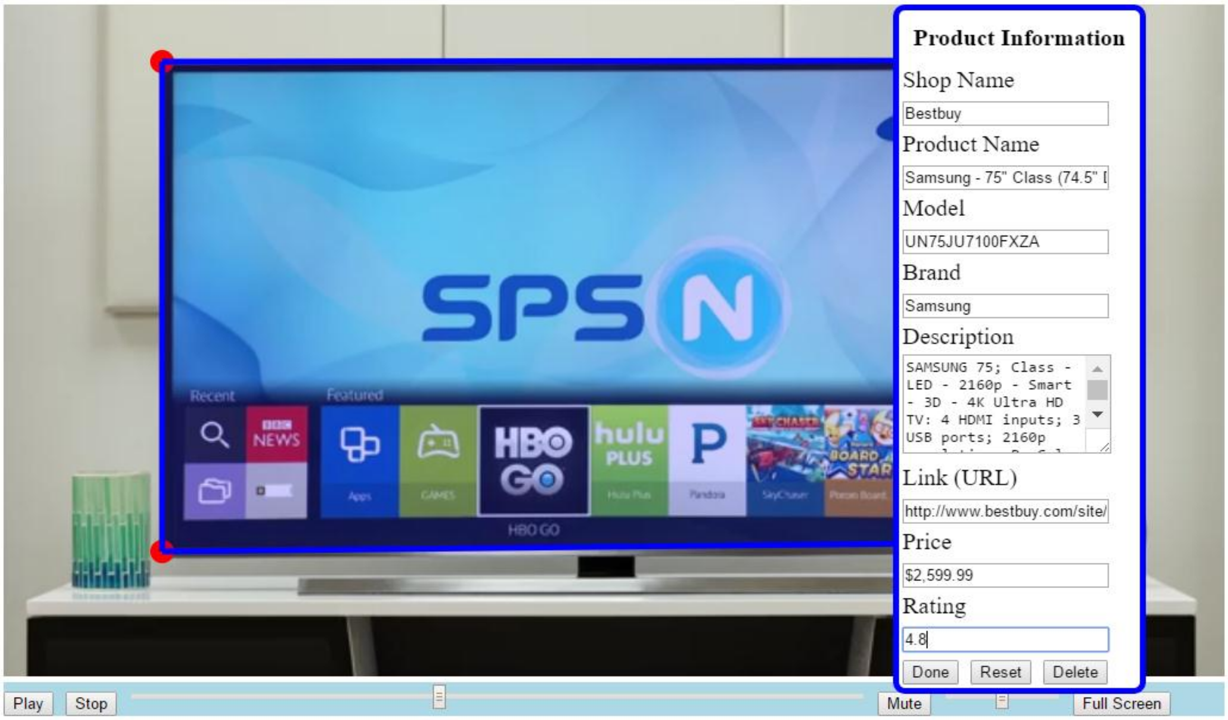
4.2Implementation

In this part, both the video annotation tool and the proposed collaborative framework are presented. Firstly, each participant must fill out the values of all of the product attributes as shown in *Figure 6*. After that, the user was allowed to view the annotated information of the chosen product by clicking the label associated with it. There was an additional feature implemented within our video annotation tool to facilitate the idea of interactive online shopping. The user could select the vendor of interest by clicking the checkbox. Subsequently, a snippet containing the information of the product offered by that vendor was displayed on the top left of the video (cf. *Figure 7*). The small piece of text also contained the link to navigate the user to the vendor's website. Using this approach, we aimed to provide a new shopping experience for the customers. The clients could watch the video and buy the product of interest simultaneously.

In the next step of the experiment, the collaborative process was separated into two phases. In the first phase, each round of contribution introduced a new participant with his/her input information for each field summarized in Figure A2 in appendices. The following steps were carried out. For each round, the average cosine distance of the whole profile $d_{t\_mean}(X)$ in (3) was re-calculated to figure out the degree of consensus within the current profile. In addition, the consensus denoted as $d_{min}(X)$ in (5) was also determined to compare with $d_{t\_mean}(X)$ to verify the condition for *susceptible to consensus*. The first phase ended when the condition in (6) was satisfied.

In this experiment, it was found that after nine rounds of contribution, the consensus within the given profile existed. The values of both $d_{t\_mean}(X)$ and $d_{min}(X)$ were recorded and plotted in *Figure 8*. When the consensus was reached among all of the participants, the quality of the collaborative process was found to be 0.397 based on (7). The quality was relatively low due to the high complexity in the vendors' ways of expression as well as the large variety of products they were willing to sell. More importantly, there was no constraint on the annotation process to guide their vendors to input the information in a similar manner.

[7] Organization of Schemas of schema.org https://schema.org/docs/schemas.html. Accessed 15 January 2017.



Product Information

Shop Name

Bestbuy

Product Name

Samsung - 75" Class (74.5" I

Model

UN75JU7100FXZA

Brand

Samsung

Description

SAMSUNG 75; Class - LED - 2160p - Smart - 3D - 4K Ultra HD TV: 4 HDMI inputs; 3 USB ports; 2160p

Link (URL)

http://www.bestbuy.com/site/

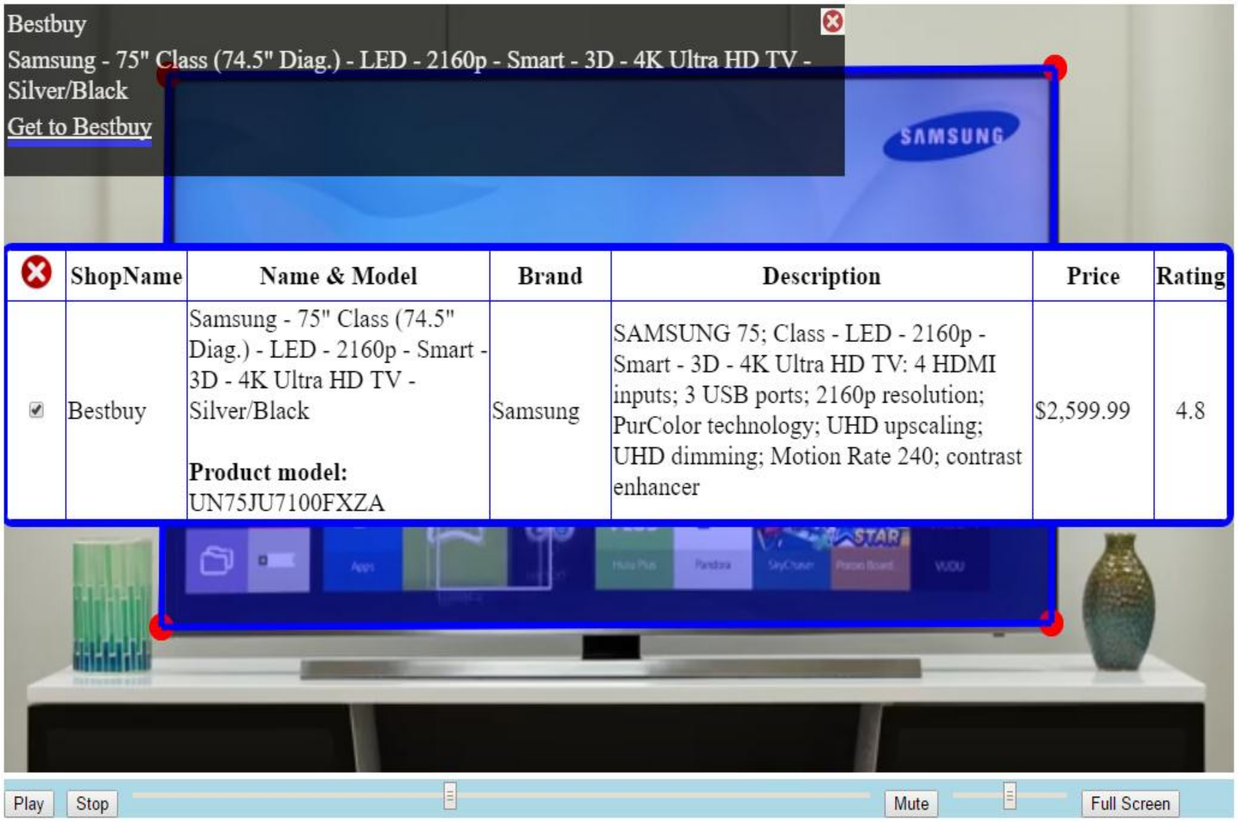
Price

$2,599.99

Rating

4.8

Figure 6 The attributes based on the vocabularies of schema.org for the product annotation supported by our system



Bestbuy

Samsung - 75" Class (74.5" Diag.) - LED - 2160p - Smart - 3D - 4K Ultra HD TV - Silver/Black

Get to Bestbuy

☒	ShopName	Name & Model	Brand	Description	Price	Rating
☑	Bestbuy	Samsung - 75" Class (74.5" Diag.) - LED - 2160p - Smart - 3D - 4K Ultra HD TV - Silver/Black **Product model:** UN75JU7100FXZA	Samsung	SAMSUNG 75; Class - LED - 2160p - Smart - 3D - 4K Ultra HD TV: 4 HDMI inputs; 3 USB ports; 2160p resolution; PurColor technology; UHD upscaling; UHD dimming; Motion Rate 240; contrast enhancer	$2,599.99	4.8

Figure 7 The information input by participant in the first round of the collaborative process displayed by our system

Round	$d_{t\_mean}(X)$	$d_{min}(X)$
1	0.000	0.000
2	0.194	0.291
3	0.362	0.437
4	0.452	0.504
5	0.498	0.547
6	0.536	0.556
7	0.558	0.573
8	0.584	0.585
9	0.610	0.603

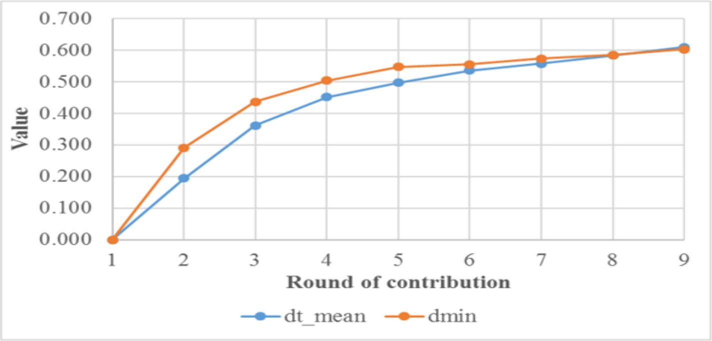
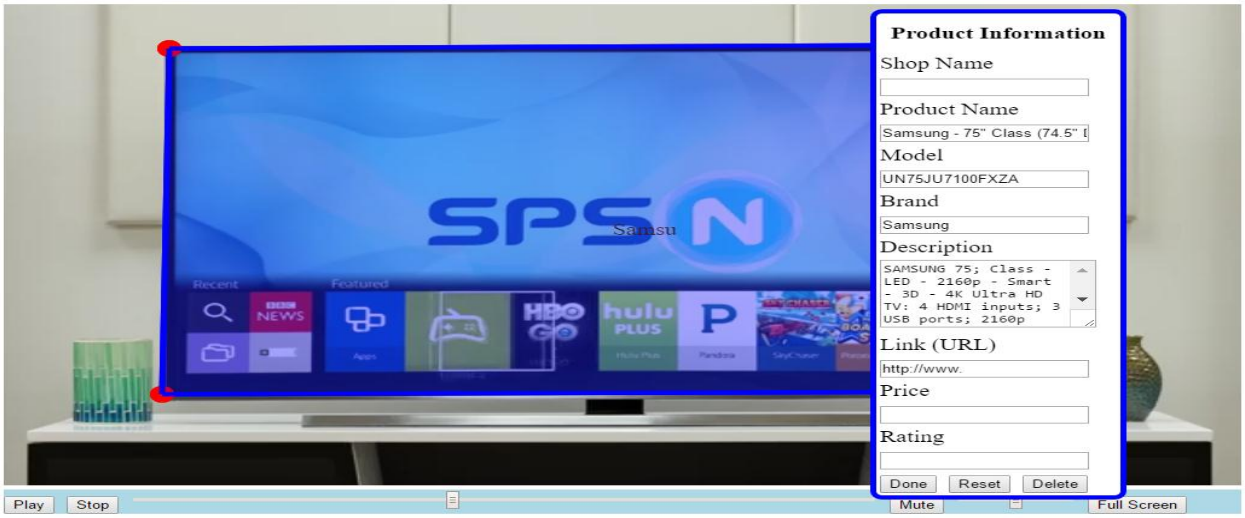


Figure 8 Values of $d_{t\_mean}(X)$ and $d_{min}(X)$ after nine rounds of contribution until the consensus condition was satisfied with the quality of 0.397

The second phase of the experiment starting from the tenth round of contribution, which was used to demonstrate the performance and robustness of our collaborative framework. At this point, we applied two different approaches. In the first scenario, after the consensus was found, we offered an auto-fill suggestion feature for the common fields (i.e., product name, model, description, and brand) using the knowledge of consensus known as from previous rounds (cf. *Figure 9*). With this support, our video annotation tool encouraged new participants to follow the closest information to all existing versions from previous rounds of contribution. Thanks to this functionality, the vendors could save their annotation efforts and input more relevant information for the customers.



Product Information

Shop Name

Product Name
Samsung - 75" Class (74.5" I

Model
UN75JU7100FXZA

Brand
Samsung

Description
SAMSUNG 75; Class - LED - 2160p - Smart - 3D - 4K Ultra HD TV: 4 HDMI inputs; 3 USB ports; 2160p

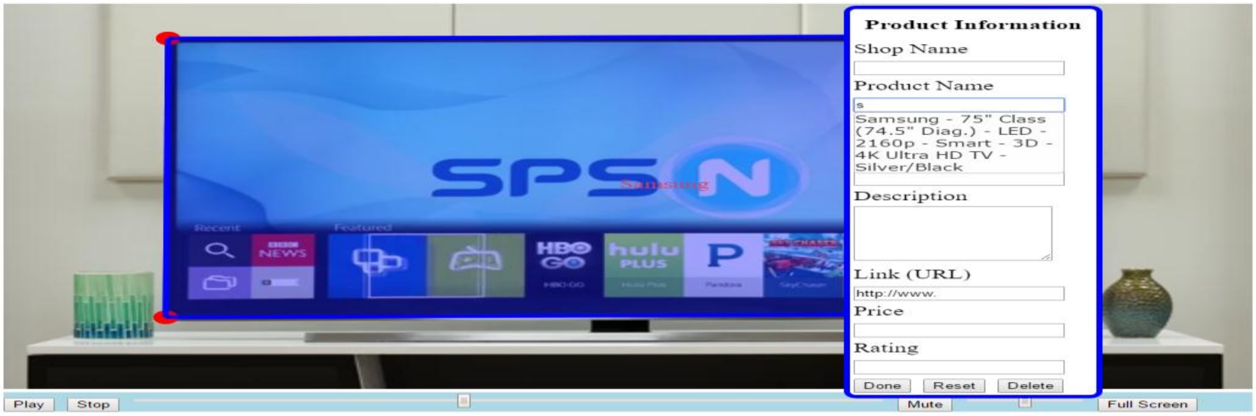
Link (URL)
http://www.

Price

Rating

Figure 9 Auto-fill suggestion feature for the tenth participant in the collaborative process using the consensus knowledge of the first participant

In contrast, we did not provide the user with such automatic recommendation in the second scenario. In other words, we let them input the annotation information without any hints just as before the consensus within the profile was achieved. To be more detailed, we invited two sets of people for each aforementioned scenario to annotate the same object. The invitation process continued until the consensus quality of either case exceeded 0.7 – the threshold to establish the credibility of the consensus. At such moment, the information of that object would be put into our database for future recommendation in case a new annotation was created (cf. *Figure 10*). After the consensus quality was sufficient to make a good compromise, we compared the rate of convergence of the consensus quality in both cases (cf. *Figure 11*).



Product Information

Shop Name

Product Name

s

Samsung - 75" Class (74.5" Diag.) - LED - 2160p - Smart - 3D - 4K Ultra HD TV - Silver/Black

Description

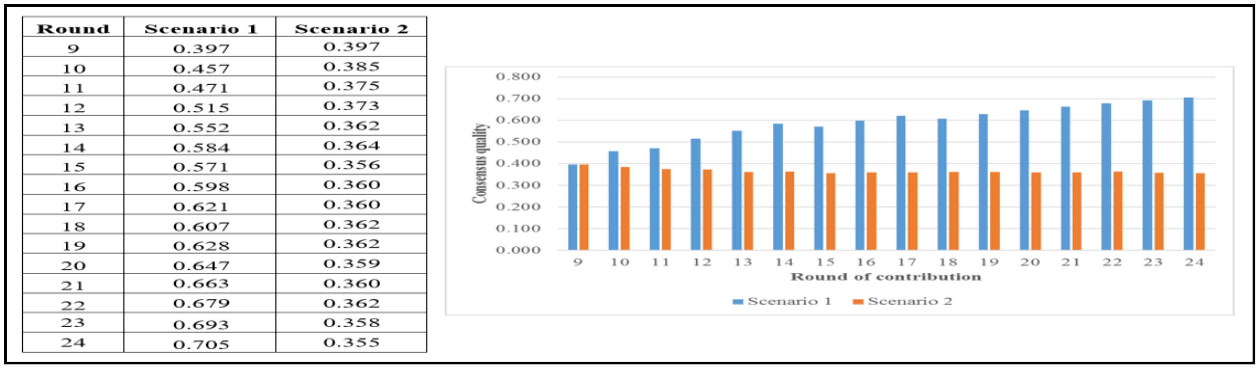
Link (URL)

http://www.

Price

Rating

Figure 10 Auto-complete recommendation feature for the product name using the previous consensus in database



Round	Scenario 1	Scenario 2
9	0.397	0.397
10	0.457	0.385
11	0.471	0.375
12	0.515	0.373
13	0.552	0.362
14	0.584	0.364
15	0.571	0.356
16	0.598	0.360
17	0.621	0.360
18	0.607	0.362
19	0.628	0.362
20	0.647	0.359
21	0.663	0.360
22	0.679	0.362
23	0.693	0.358
24	0.705	0.355

Figure 11 Consensus quality comparison between ***with*** and ***without*** the use of consensus knowledge for suggestion during the annotation process

From the chart, we could see that *scenario 1* was better than *scenario 2* in terms of the consensus quality (i.e., 98.6% higher). According to [23], the *profile 1* had a higher probability of forming a good compromise than that of the *profile 2*. More specifically, after 24 rounds, thanks to the aid of consensus knowledge, the profile was sufficiently reliable to become the knowledge of the current object for both clients and vendors. It was also worth noting that the sporadic decreases in the quality of the *profile 1* in the 15th and 18th rounds were due to the flexibility of our system. Under such circumstances, the participants could alter the recommended information to fit their selling strategies. On the contrary, the participants of the *profile 2* were given no clue about the common knowledge of the previous contribution. They tended to input their own opinions for the product attributes. Consequently, the consensus quality was fluctuating without any clear convergence. In brief, our collaborative framework significantly increased the relevance of the content displayed to the customers, which in turn helped them make the most appropriate purchase decisions.

5.Conclusion and future work

In this study, we have proposed a collaborative framework for the user to create annotations on the video using the vocabularies provided by the GoodRelations ontology. The system also supports the collaborative process using the consensus quality metrics. Our solution has shown to provide an alternative approach to other traditional hypervideo tools. To put it another way, the system helps balance between product identity and product offering with the aid of two different types of product attributes. More specifically, the identity of a specific product can be ensured based consensus knowledge of the product name, model, brand and description; whereas, the product offering is characterized by the vendor’s information and their offered price and rating. With this approach, the buyers can easily identify their product of interest to make better purchase decisions, and the suppliers know the item to make appropriate advertisement. With this information, the vendors can make various offers with different levels of benefits for the same product for the clients to select freely. As a result, the quality of online shopping can be improved dramatically.

In the future, we plan to conduct further research to test the system on a larger scale with more attributes of the product and different sources of video. Moreover, it is also promising to apply modern natural language processing technologies to reduce the confusion caused by the diversity in the way the users create their annotations. By this way, our framework can achieve the consensus much faster. In overall, we can improve the robustness of both the annotation tool and the collaborative framework.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Gomez-Perez A, Fernández-López M, Corcho O. Ontological Engineering: with examples from the areas of knowledge management, e-Commerce and the semantic web. Springer Science and Business Media; 2006.

[2] Hepp M. Goodrelations: An ontology for describing products and services offers on the web. In international conference on knowledge engineering and knowledge management 2008 (pp. 329-46). Springer Berlin Heidelberg.

[3] Duong T, Rosli A, Sean V, Lee KS, Jo GS. E-Commerce video annotation using good relations-based LODs with faceted search in smart TV environment. International conference on computational collective intelligence 2012 (pp.253-63).

[4] Lippman A. Movie-maps: an application of the optical videodisc to computer graphics. In ACM siggraph computer graphics 1980 (pp. 32-42). ACM.

[5] Aubert O, Prié Y. Advene: active reading through hypervideo. In proceedings of the sixteenth ACM conference on hypertext and hypermedia 2005 (pp. 235-44). ACM.

[6] Jeong JW, Hong HK, Lee DH. Ontology-based automatic video annotation technique in smart TV environment. IEEE Transactions on Consumer Electronics. 2011;57(4).

[7] Naphade MR, Smith JR. On the detection of semantic concepts at TRECVID. In proceedings of the 12th annual ACM international conference on Multimedia 2004 (pp. 660-7). ACM.

[8] Park KW, Lee JH, Moon YS, Park SH, Lee DH. OLYVIA: ontology-based automatic video annotation and summarization system using semantic inference rules. In international conference on semantics, knowledge and grid 2007 (pp. 170-5). IEEE.

[9] Vondrick C, Patterson D, Ramanan D. Efficiently scaling up crowdsourced video annotation. International Journal of Computer Vision. 2013; 101(1):184-204.

[10] Yuen J, Russell B, Liu C, Torralba A. Labelme video: Building a video database with human annotations. In international conference on computer vision 2009 (pp. 1451-8). IEEE.

[11] Lee KS, Rosli AN, Supandi IA, Jo GS. Dynamic sampling-based interpolation algorithm for representation of clickable moving object in collaborative video annotation. Neurocomputing. 2014; 146:291-300.

[12] Khusro S, Khan M, Ullah I. Collaborative video annotation based on ontological themes, temporal duration and pointing regions. In proceedings of the international conference on informatics and systems 2016 (pp. 121-6). ACM.

[13] Goy A, Magro D, Petrone G, Picardi C, Segnan M. Ontology-driven collaborative annotation in shared workspaces. Future Generation Computer Systems. 2016; 54:435-49.

[14] Carbonaro A, Ferrini R. Ontology-based video annotation in multimedia entertainment. Consumer communications and networking conference 2007(pp. 1087-91).

[15] Xue M, Zheng S, Zhang C. Ontology-based surveillance video archive and retrieval system. In fifth international conference on advanced computational intelligence 2012 (pp. 84-9). IEEE.

[16] Ouyang JQ, Jin-tao L, Yong-dong Z. Ontology based sports video annotation and summary. In content computing 2004 (pp. 499-508). Springer Berlin Heidelberg.

[17] Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a research tool for general practice?. Family Practice. 1993; 10(1):76-81.

[18] Pill J. The Delphi method: substance, context, a critique and an annotated bibliography. Socio-Economic Planning Sciences. 1971; 5(1):57-71.

[19] Martinez-Cruz C, Blanco IJ, Vila MA. Ontologies versus relational databases: are they so different? A comparison. Artificial Intelligence Review. 2012; 38(4):271-90.

[20] Shimrat M. Algorithm 112: position of point relative to polygon. Communications of the ACM. 1962; 5(8):434.

[21] Duong TH, Nguyen NT, Nguyen DC, Nguyen TP, Selamat A. Trust-based consensus for collaborative ontology building. Cybernetics and Systems. 2014; 45(2):146-64.

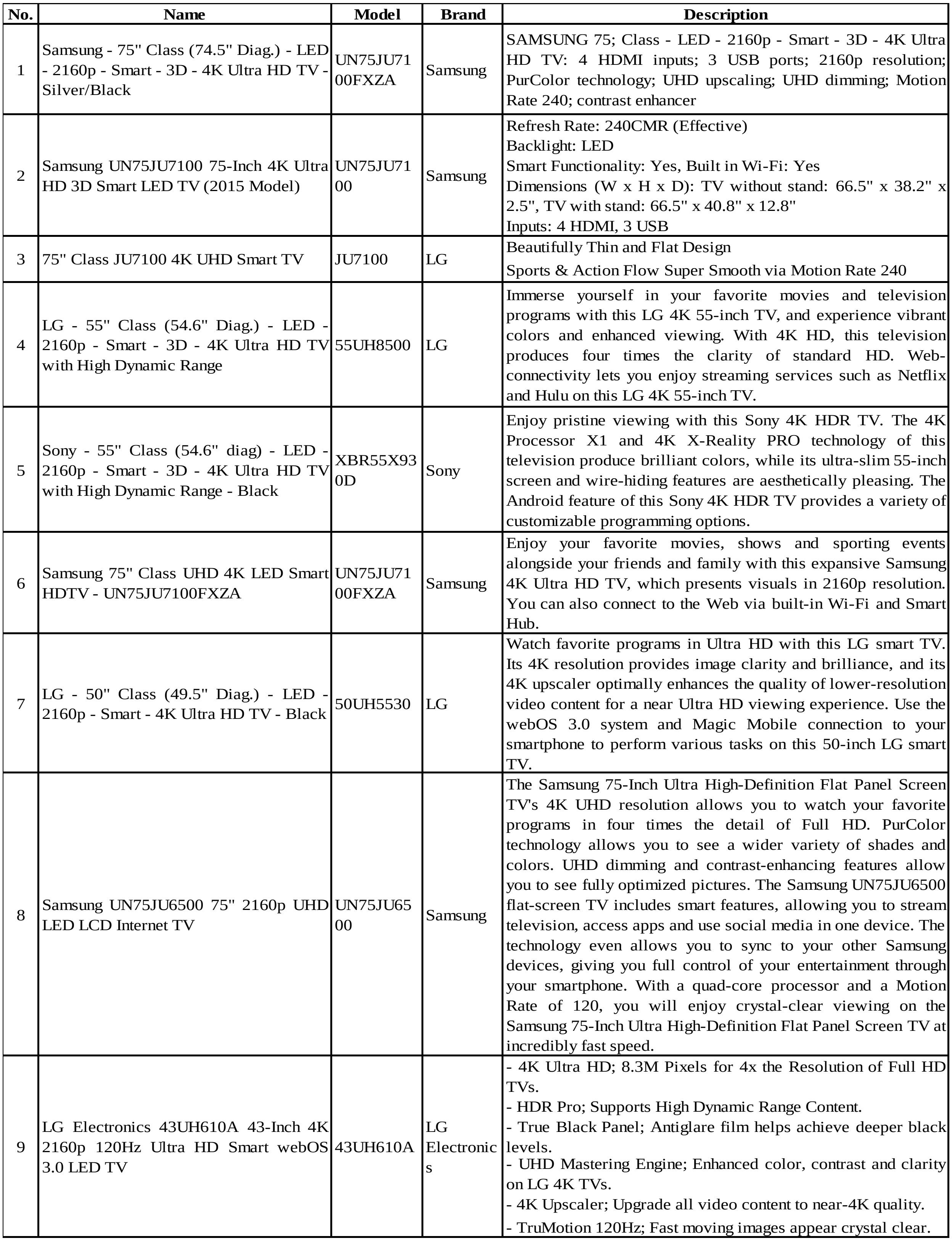
[22] Duong TH, Tran MQ, Nguyen TP. Collaborative Vietnamese WordNet building using consensus quality. Vietnam Journal of Computer Science. 2017;4(2):85-96.

[23] Nguyen NT. Advanced methods for inconsistent knowledge management. Springer Science & Business Media; 2007.

Appendix

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="objects">                                 <!-- The root element -->
   <xs:complexType>
     <xs:sequence>
      <xs:element type="xs:string" name="vid"/>               <!-- Unique id of the current video -->
      <xs:element name="polygon">                             <!-- A single annotation -->
        <xs:complexType>
         <xs:sequence>
           <xs:element type="xs:string" name="name"/>         <!-- Name of the current annotation -->
           <xs:element type="xs:string" name="creator"/>      <!-- Creator of the current annotation -->
           <xs:element type="xs:float" name="startFrame"/>   <!-- Frame when the current annotation appears -->
           <xs:element type="xs:float" name="endFrame"/>     <!-- Frame when the current annotation disappears -->
           <xs:element name="shopList">                       <!-- List of offerings for the current annotation -->
            <xs:complexType>
              <xs:sequence>
               <xs:element name="shop">
                 <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:string" name="shopCreator"/>   <!-- Creator of the current shop -->
                    <xs:element type="xs:string" name="shopName"/>      <!-- Name of the current shop -->
                    <xs:element type="xs:string" name="productName"/> <!-- Product name offered by the current shop-->
                    <xs:element type="xs:string" name="model"/>         <!-- Offered product model -->
                    <xs:element type="xs:string" name="brand"/>         <!-- Offered product brand -->
                    <xs:element type="xs:string" name="description"/>   <!-- Offered product description -->
                    <xs:element type="xs:string" name="link"/>          <!-- Link of the website of the shop -->
                    <xs:element type="xs:float" name="price"/>          <!-- Price of the offered product -->
                    <xs:element type="xs:float" name="rate"/>           <!-- Rating of the offered product -->
                  </xs:sequence>
                 </xs:complexType>
               </xs:element>
              </xs:sequence>
            </xs:complexType>
           </xs:element>
           <xs:element type="xs:integer" name="nPoints"/>          <!—No. of points of the polygon-like annotation -->
           <xs:element name="keyFrames">                           <!-- Keyframes of used for linear interpolation -->
            <xs:complexType>
              <xs:sequence>
               <xs:element type="xs:float" name="frame">           <!-- Specific keyframe of the current annotation -->
              </xs:sequence>
            </xs:complexType>
           </xs:element>
           <xs:element name="pointsChange">                        <!-- The positions of the above keyframes -->
            <xs:complexType>
              <xs:sequence>
               <xs:element name="order">                           <!-- Position of the annotation at a keyframe -->
                 <xs:complexType>
                  <xs:sequence>
                    <xs:element name="point">                      <!-- Coordinate of a point of the current polygon -->
                     <xs:complexType>
                       <xs:sequence>
                        <xs:element type="xs:float" name="x"/>     <!-- x-coordinate of the current point -->
                        <xs:element type="xs:float" name="y"/>     <!-- y-coordinate of the current point -->
                       </xs:sequence>
                     </xs:complexType>
                    </xs:element>
                  </xs:sequence>
                 </xs:complexType>
               </xs:element>
         </xs:sequence>
        </xs:complexType>
       </xs:element>
       <xs:element name="hideFrames">              <!-- Frames when the current annotation is hidden -->
        <xs:complexType>
         <xs:sequence>
          <xs:element name="frame">                  <!-- Hidden frames specified by the starting and ending frames -->
           <xs:complexType>
            <xs:sequence>
             <xs:element type="xs:float" name="s"/>          <!-- Starting hidden frame -->
             <xs:element type="xs:float" name="d"/>          <!-- Ending hidden frame -->
            </xs:sequence>
           </xs:complexType>
          </xs:element>
         </xs:sequence>
        </xs:complexType>
       </xs:element>
       <xs:element name="staticFrames">           <!-- Frames when the current annotation does not move -->
        <xs:complexType>
         <xs:sequence>
          <xs:element name="frame">                 <!—A static period specified by the starting and ending frames -->
           <xs:complexType>
            <xs:sequence>
             <xs:element type="xs:float" name="s"/>          <!-- Starting static frame -->
             <xs:element type="xs:float" name="d"/>          <!-- Ending static frame -->
            </xs:sequence>
           </xs:complexType>
          </xs:element>
         </xs:sequence>
        </xs:complexType>
       </xs:element>
       <xs:element name="consensus">                                <!-- Consensus of the current annotated product -->
        <xs:complexType>
          <xs:sequence>
          <xs:element type="xs:string" name="productName"/>  <!-- Product name of the consensus -->
          <xs:element type="xs:string" name="model"/>              <!-- Product model of the consensus -->
          <xs:element type="xs:string" name="brand"/>              <!-- Product brand of the consensus -->
          <xs:element type="xs:string" name="description"/>       <!-- Product description of the consensus -->
          <xs:element type="xs:float" name="quality"/>              <!—Consensus quality in the range [0 - 1] -->
         </xs:sequence>
        </xs:complexType>
       </xs:element>
      </xs:sequence>
     </xs:complexType>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

Figure A1 XML schema definition of the XML-based metadata file for video annotation used in our framework in which the meaning of each element is put inside the ″<!-- -->" symbol



No.	Name	Model	Brand	Description
1	Samsung - 75" Class (74.5" Diag.) - LED - 2160p - Smart - 3D - 4K Ultra HD TV - Silver/Black	UN75JU7100FXZA	Samsung	SAMSUNG 75; Class - LED - 2160p - Smart - 3D - 4K Ultra HD TV: 4 HDMI inputs; 3 USB ports; 2160p resolution; PurColor technology; UHD upscaling; UHD dimming; Motion Rate 240; contrast enhancer
2	Samsung UN75JU7100 75-Inch 4K Ultra HD 3D Smart LED TV (2015 Model)	UN75JU7100	Samsung	Refresh Rate: 240CMR (Effective) Backlight: LED Smart Functionality: Yes, Built in Wi-Fi: Yes Dimensions (W x H x D): TV without stand: 66.5" x 38.2" x 2.5", TV with stand: 66.5" x 40.8" x 12.8" Inputs: 4 HDMI, 3 USB
3	75" Class JU7100 4K UHD Smart TV	JU7100	LG	Beautifully Thin and Flat Design Sports & Action Flow Super Smooth via Motion Rate 240
4	LG - 55" Class (54.6" Diag.) - LED - 2160p - Smart - 3D - 4K Ultra HD TV with High Dynamic Range	55UH8500	LG	Immerse yourself in your favorite movies and television programs with this LG 4K 55-inch TV, and experience vibrant colors and enhanced viewing. With 4K HD, this television produces four times the clarity of standard HD. Web-connectivity lets you enjoy streaming services such as Netflix and Hulu on this LG 4K 55-inch TV.
5	Sony - 55" Class (54.6" diag) - LED - 2160p - Smart - 3D - 4K Ultra HD TV with High Dynamic Range - Black	XBR55X930D	Sony	Enjoy pristine viewing with this Sony 4K HDR TV. The 4K Processor X1 and 4K X-Reality PRO technology of this television produce brilliant colors, while its ultra-slim 55-inch screen and wire-hiding features are aesthetically pleasing. The Android feature of this Sony 4K HDR TV provides a variety of customizable programming options.
6	Samsung 75" Class UHD 4K LED Smart HDTV - UN75JU7100FXZA	UN75JU7100FXZA	Samsung	Enjoy your favorite movies, shows and sporting events alongside your friends and family with this expansive Samsung 4K Ultra HD TV, which presents visuals in 2160p resolution. You can also connect to the Web via built-in Wi-Fi and Smart Hub.
7	LG - 50" Class (49.5" Diag.) - LED - 2160p - Smart - 4K Ultra HD TV - Black	50UH5530	LG	Watch favorite programs in Ultra HD with this LG smart TV. Its 4K resolution provides image clarity and brilliance, and its 4K upscaler optimally enhances the quality of lower-resolution video content for a near Ultra HD viewing experience. Use the webOS 3.0 system and Magic Mobile connection to your smartphone to perform various tasks on this 50-inch LG smart TV.
8	Samsung UN75JU6500 75" 2160p UHD LED LCD Internet TV	UN75JU6500	Samsung	The Samsung 75-Inch Ultra High-Definition Flat Panel Screen TV's 4K UHD resolution allows you to watch your favorite programs in four times the detail of Full HD. PurColor technology allows you to see a wider variety of shades and colors. UHD dimming and contrast-enhancing features allow you to see fully optimized pictures. The Samsung UN75JU6500 flat-screen TV includes smart features, allowing you to stream television, access apps and use social media in one device. The technology even allows you to sync to your other Samsung devices, giving you full control of your entertainment through your smartphone. With a quad-core processor and a Motion Rate of 120, you will enjoy crystal-clear viewing on the Samsung 75-Inch Ultra High-Definition Flat Panel Screen TV at incredibly fast speed.
9	LG Electronics 43UH610A 43-Inch 4K 2160p 120Hz Ultra HD Smart webOS 3.0 LED TV	43UH610A	LG Electronics	- 4K Ultra HD; 8.3M Pixels for 4x the Resolution of Full HD TVs. - HDR Pro; Supports High Dynamic Range Content. - True Black Panel; Antiglare film helps achieve deeper black levels. - UHD Mastering Engine; Enhanced color, contrast and clarity on LG 4K TVs. - 4K Upscaler; Upgrade all video content to near-4K quality. - TruMotion 120Hz; Fast moving images appear crystal clear.

Figure A2 Annotation information input by the participants in the nine first rounds of contribution

Triet Huynh Minh Le is currently a senior student with a major in Computer Science at International University-Vietnam National University HCMC. His research interest focuses on semantic video annotation with consensus-based collaboration, data mining, and machine learning.

Trong Hai Duong got M.Sc. and Ph.D. degree at school of Computer and Information Engineering, Inha University, 2012. His research interest focuses on ontology and semantic web, big data, smart data, and ecommerce systems. He has more than 40 publications. Currently, he works for International University- Vietnam National University HCMC.
Email: haiduongtrong@gmail.com