Complura: Exploring and Leveraging a Large-scale Multilingual Visual Sentiment Ontology

Hongyi Liu*†, Brendan Jou*†, Tao Chen*†, Mercan Topkara*‡, Nikolaos Pappas*§, Miriam Redi*¶, Shih-Fu Chang†

†Columbia University, New York, NY USA
‡JW Player, New York, NY USA
§Idiap Research Institute/EPFL, Martigny, Switzerland
¶Yahoo, London, UK

ABSTRACT

What would someone from another culture think of this photograph I just took? Would they think my picture of this 'wilted flower' was also sentimentally positive or would they perceive it negatively instead? Or what if I wanted to find other photographs that are semantically related to my image as well as sentimentally sensitive, but from other cultures? In fact, this cultural and sentimental relevancy are features that we would expect of any recommender system and query expansion engine, respectively. Motivated by this, we present an online demonstration of a system called **Complura**. Our system implements three major functions: an interactive multilingual ontology browser, a cross-lingual image-based sentiment analyzer, and a culturally-coherent, sentiment-aware image query expansion engine. We ground our system on a multilingual visual sentiment ontology, containing over 10k sentiment-polarized visual concepts over 12 languages and over 7.3M images.

CCS Concepts

•Information systems → Multimedia databases; Recommender systems; Sentiment analysis; •Human-centered computing → Information visualization; Visualization toolkits;

Keywords

Multilingual; Culture; Ontology; Browsing Interfaces; Sentiment Analysis; Query Expansion

1. INTRODUCTION

The uniqueness and intertwining of culture, language and sentiment are some of the things that make our world so diverse and beautiful. Amidst the tremendous amount of social multimedia content shared around world each day, how images in various languages are used to express language-specific concepts, as well as how these concepts connect to emotion and sentiment remain open questions. Recently, a large-scale multilingual visual sentiment ontology (MVSO), consisting of over 15,000 sentiment- and emotion-polarized visual bi-concepts called adjective-noun pairs (ANP) over 12 languages, was proposed for language-dependent content analysis [6]. The adjective-noun pair semantic structure used in the ontology is a powerful construct where the noun component defines a visually grounded, and hopefully machine-detectable object or concept, and the adjective contributes a sentiment or emotion attachment. This multilingual visual sentiment ontology provides researchers a rich information source for analyzing cultural connections and characteristics of visual sentiment across language lines. In

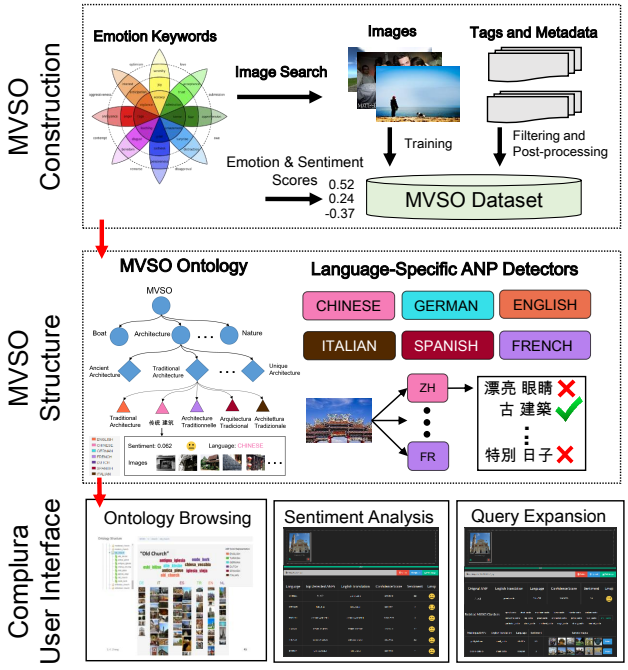

Figure 1: *Complura* System Architecture Overview. The Multilingual Visual Sentiment Ontology [6] (MVSO) forms the basis of *Complura*'s backend.

*Denotes equal contribution.

ICMR'16, June 06-09, 2016, New York, NY, USA

DOI: http://dx.doi.org/10.1145/2911996.2912030

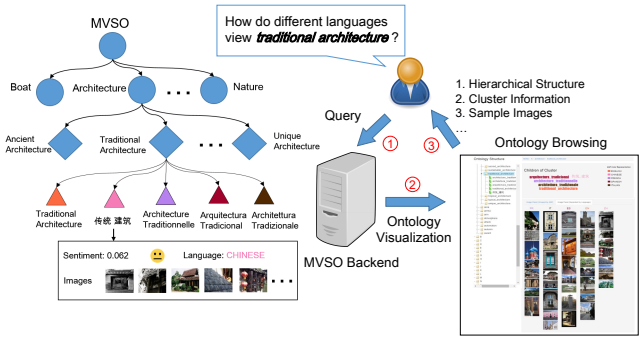

Figure 2: Workflow of *Complura*'s Ontology Browser. A researcher with a question or hypothesis about how visual affective concepts are understood across languages or sentimentally portrayed can explore the MVSO corpus by querying Complura and exploring the ontology. *Complura* visualizes the hierarchical structure, cluster information and shows related sample images in mined from Flickr to help users to answer their research questions about multilingual visual sentiment.

addition, it carries the potential for novel augmentations to retrieval and recommendation tasks in social media.

In order to explore hierarchically structured information, a number of systems have been proposed in the past [2]. For example, a node-link paradigm is utilized by many systems [9, 5, 10] for visualization, while others like *CropCircles* [14] use concentric circles to represent hierarchical ontology. And some hybrid works like *Knoocks* [7], *OntoTrix* [1] and *NodeTrix* [4] combine node-link, nested blocks and adjacency matrix representations. These systems provide deep insights into the visualization of their respective ontologies, but they seldom provide integrated functions that allow users to leverage the ontology for other purposes.

In this work, we aim to answer the question of how to understand and leverage the cultural influence given the massive number of images shared in social media expressing strong sentiments and emotions. Our prior work [6] focused on the construction and analysis of the MVSO dataset, but it did not provide an intuitive system for individuals to explore and leverage this ontology. We present an integrated system called **Complura**[1], that allows users to formulate and verify hypothesis by interactively browsing and exploring the ontology, examine the image sentiment of different languages, and perform cultural-coherently expanded image query through leveraging the MVSO dataset.

Our proposed system provides three major functions: (1) an interactive ontology browsing interface for a multilingual visual sentiment ontology; (2) a language-specific image-based sentiment analyzer; and (3) a culturally-coherent image query expansion engine with sentiment-aware sensitivity. To the best of our knowledge, **Complura** is the first system to explore and utilize the hierarchical structure of such a large-scale multilingual visual sentiment ontology.

2. ONTOLOGY ORGANIZATION

We develop **Complura** on top the Multilingual Visual Sentiment Ontology (MVSO) [6] and its accompanying image dataset because of its rich structured cross-cultural information with visual sentiment in different languages. This makes it possible for us to both evaluate the sentiment of visual content and create novel applications in a multicultural context. The development of the MVSO in [6] can be summarized in two major stages: a data aggregation and a ontology construction stage.

Table 1: Summary Statistics of the Multilingual Visual Sentiment Ontology (MVSO) Dataset [6].

#Languages	12
Total #ANPs	15,630
Total #Images	7,368,364
Average #Images/ANP	471

2.1 Data Aggregation

The data aggregation stage of MVSO dataset begins with a set of emotion seed keywords selected from Plutchik's Wheel of Emotions [12]. These seed keywords, inspired by psychology theory, including *joy*, *trust*, *fear*, and so on, are translated by native speakers of 12 different languages: Arabic, Chinese, Dutch, English, French, German, Italian, Persian, Polish, Russian, Spanish and Turkish. These keywords are used to query for images along with their tags and metadata on the social multimedia platform, Flickr[2]. Part-of-speech labelling and adjective-noun combinations are performed to generate semantic structures called adjective-noun pairs (ANPs), such as *young children*, *smiling face* and *nice car*. The final ANP set is obtained through a course-to-fine filtering process on candidate ANPs based on language, semantics, sentiment, frequency and diversity. Sentiment and emotion scores are also calculated and crowdsourced for each ANP. A set of 15,630 ANPs are ultimately used to also create a multilingual image dataset of over 7.3M social images. And the images are used in [6] to train convolutional neural network (CNN)-based, language-specific visual concept detectors of ANPs [8] on a subset of six major languages cover over 10k of the ANPs. The classification performance of the ANP detectors range from ∼10.1% for English on 4,342 ANPs to ∼30.1% for German on 275 ANPs. A more detailed explanation can be found in [6].

2.2 Ontology Construction

The construction of hierarchically structured ANPs in the MVSO follows a bottom-up strategy, including exact matching, noun-based grouping and synonym-based group merging. In particular, we first group ANPs of different languages that share the same English translation together. For example, 传统 建筑 (ZH), *architecture traditionnelle* (FR), *architettura tradizionale* (ES), *traditional architecture* (EN) are grouped together as they all have the same English translation (*traditional architecture*). Then all the groups, or clusters, represented by English translation of ANPs, are further grouped by noun. English translations like *ancient architecture*, *beautiful architecture*, *traditional architecture* are grouped as they share the same noun (*architecture*). And finally, noun groups that are synonyms, such as *clothes* and *dress* are merged together to enhance the compactness and diversity of the ontology.

[1]http://mvso.cs.columbia.edu/complura

[2]https://www.flickr.com

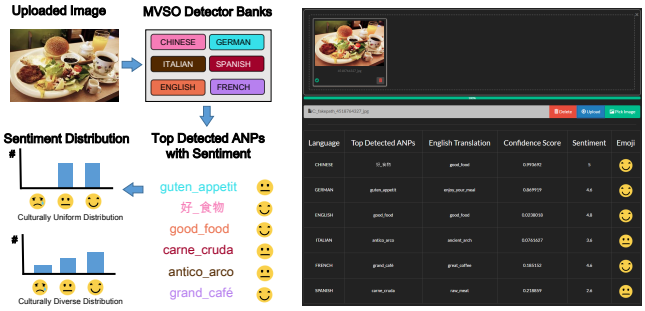

Figure 3: *Left*: Three-stage workflow of image sentiment analysis across languages. Given an input image, top detected ANPs of different languages with corresponding sentiment scores are returned. *Right*: Screenshot of interface for image sentiment analysis.

3. SYSTEM OVERVIEW

As shown in Figure 1, our system **Complura** builds on two main aspects of MVSO: the ontology and language-specific ANP detectors. The ontology refers to the hierarchical structure of the ANPs, and is valuable for capturing mid-level, sentiment-biased semantic information across languages, where each ANP is associated with a sentiment value and related images. There are also six ANP detector banks trained in [6] capable of detecting the presence of ANPs from Chinese, German, English, Italian, Spanish and French in images. The goal of **Complura** is to assist users in exploring and leveraging the ontology for semantic analysis of visual content across multiple languages. We implement three major features: ontology browsing, sentiment analysis and multilingual image-based query expansion.

3.1 Ontology Browsing

We focus on high interaction browsing of the Multilingual Visual Sentiment Ontology in **Complura**. The ontology browser enables users to formulate hypothesis, choose controlled categories, and visualize image differences tagged with different languages. Given a scenario like the one shown in Figure 2, if a user wants to research the cultural difference of the visual concept *traditional architecture*, they need only use the ontology browser in our system which visualizes MVSO in a succinct, intuitive and interactive browser interface. In the web interface, as shown in the bottom-right of Figure 2, the ontology browser consists of three main panels: the ontology-structure panel (left), word cloud panel (top-right) and image comparison panel (bottom-right).

The hierarchical structure is visualized using an interactive tree library[3], which provides folder-like expansion interactions for the user to navigate through the ontology. This approach offers a top-down view of the whole ontology. For non-leaf nodes in the ontology, the word cloud panel allows user to intuitively grasp the content of certain node. When a user mouses over a word cloud item, a hover box is shown and displays additional contextual information for that item like English translations and number of child nodes. As shown in the Figure 2, each column of the visualization page (right side) show a maximum of eight images related to specific ANP, and a language may occupy multiple columns, depending on the number of ANPs that appear in this group. All the ANPs in a group share the same English translation, which generally corresponds to their semantic coherence. This column-based panel allows users to compare both inter- and intra-language connections and discrepancies of a certain ANP, all at the same time.

Consider again the scenario where a researcher wants to explore the cross-cultural differences of a visual concept like *traditional architecture*, they can directly compare the visual difference of sample images of *traditional architecture* between different languages on a single page in **Complura**. In addition, they can easily navigate to the siblings of the target node, such as *ancient architecture*, *modern architecture* and so on. Without this ontology visualizer, users may have had to manually find the translations of *traditional architecture* in different languages, and then perform text- or tag-based image query on search engines like Google Images or Flickr, which is tedious and inconvenient.

[3]https://www.jstree.com

3.2 Ontology Applications

Here, we utilize the unique ontological structure and the language-specific ANP detectors of MVSO to implement two novel applications: language-specific image sentiment analysis and culturally-coherent image query expansion.

3.2.1 *Multilingual Image Sentiment Analysis*

In our ontology, each ANP is associated with a sentiment score, which is calculated using SentiStrength [13] and SentiWordnet [3], and verified by crowdsourcing validation [11]. Using these scores, we seek to perform multilingual sentiment analysis for a given image using the MVSO ANP detector banks [6] as a proxy.

Imagine a scenario where a Chinese businessperson is designing a website to advertise *good food* or *modern art* but wants to expand to markets of different languages. In this case, it is critical that they pick the culturally-sensitive images to use in these other markets. One solution that **Complura** provides is an application mode where users can check if an image is appropriate. A user can choose and upload an image to find if the image is suitable for foreign markets. We run our ANP detector banks from various languages on it and given the top detected ANP in each language, and the user can make a judgment on its cultural sensitivity, e.g. based on the sentiment of the top detected ANP. In **Complura**, we enable an option for the user to also find more images of the detected ANPs in foreign languages. By checking the sentiment distribution of the detected ANPs across different languages, the user can determine whether their image is uniformly perceived in sentiment across cultures. This multilingual sentiment analysis process is shown in Figure 3.

3.2.2 *Multicultural Image Query Expansion*

Another application of MVSO and feature in **Complura** is an image query expansion engine with cultural and sentiment coherence using the semantic structure and detectors in [6]. As illustrated in the Figure 4, users begin by choosing a specified ANP detector of certain language and upload an image to the server. The system then detects ANPs in specified language, and ANPs with culturally-coherent semantics are retrieved through traversing a related sub-tree in the ontology. Some results are then pruned by ANP detector confidence scores as well as sentiment scores of re-

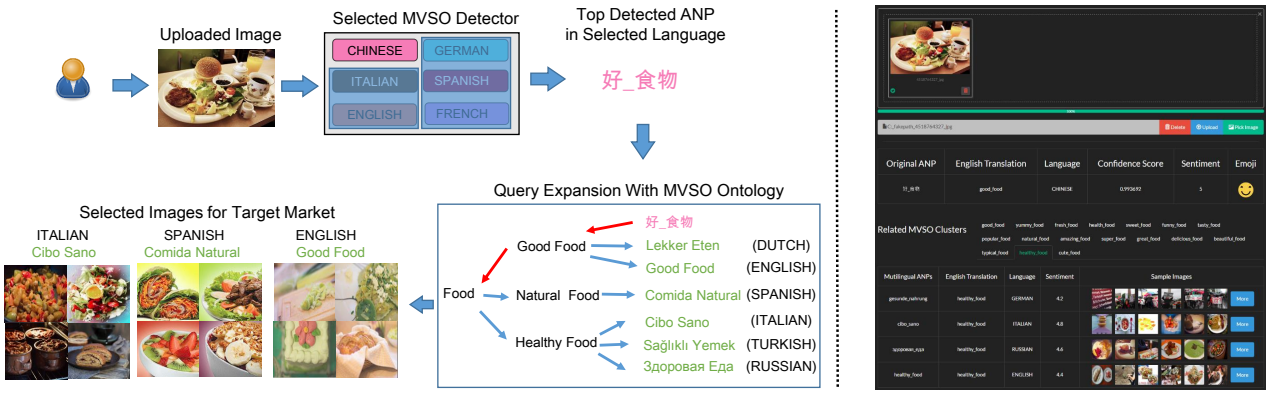

Figure 4: *Left:* Workflow of sentiment sensitive image query expansion across languages. Given an image and selected language, the top detected ANP using detectors of that language is applied toward query expansion by discovering related ANPs in the hierarchical structure of the ontology. Sample images and related ANPs are returned to user for further comparison and selection. *Right:* Screenshot of interface for query expansion.

trieved ANPs. Thus, our system is able to return ANPs of the specified language, together with other ANPs that are semantically- and sentimentally-coherent in other languages. Image query expansion is then achieved by simply returning images associated with the retrieved multilingual ANPs.

Following the same scenario in Section 3.2.1, the Chinese businessperson can also perform image query expansion using **Complura**. After uploading a target image and selecting the origin language, i.e. Chinese in this case, **Complura** will then run the Chinese ANP detectors on the image. The system maps the detected Chinese ANPs to the top MVSO clusters that elicit distinct and uniform semantics. From each of these mapped MVSO clusters, the user can then view the English translated meanings and the sentiment of the retrieved multilingual ANPs. We also show sample images related to retrieved ANPs from the other languages. Compared with the image sentiment analysis application, which focuses on checking the "correctness" of visual content across languages, **Complura**'s image query expansion aims to provide the user with images that have semantic and sentimental coherence across multiple languages. Thus, users are able to pick suitable images that are associated with relevant ANPs for each foreign market.

4. CONCLUSIONS & FUTURE WORK

We have presented a web-based system called **Complura** for exploring and leveraging a Multilingual Visual Sentiment Ontology [6]. **Complura** is built on hierarchically organized sentiment-biased visual concepts and language-specific adjective-noun pair detectors. **Complura** implements three major functions: an interactive multilingual ontology browser, a cross-lingual image-based sentiment analyzer and a culturally-coherent, sentiment-sensitive image query expansion engine. We provided corresponding usage scenarios for each function to show that it is a convenient tool that facilitates multilingual content analysis for researchers and illustrated its potential business value.

In the future, we plan to enhance our system along two major dimensions. The first is to enlarge the coverage in [6] to include more comprehensive entities, e.g. more ANPs across more languages. In addition, we want to explore other human factors such as gender and age to add more diversity.

5. REFERENCES

[1] B. Bach, G. Legostaev, and E. Pietriga. Visualizing populated ontologies with OntoTrix. In *Intl Journ. on Semantic Web & Info. Sys.*, 2010.

[2] N. Bikakis and T. Sellis. Exploration and visualization in the web of big linked data: A survey of the state of the art. *arXiv preprint arXiv:1601.08059*, 2016.

[3] A. Esuli and F. Sebastiani. SentiWordnet: A publicly available lexical resource for opinion mining. In *Language Resources and Evaluation Conf. (LREC)*, volume 6, 2006.

[4] N. Henry, J.-D. Fekete, and M. J. McGuffin. Nodetrix: A hybrid visualization of social networks. *IEEE Trans. on Vis. and Comp. Graphics*, 13(6), 2007.

[5] W. Hop, S. de Ridder, F. Frasincar, and F. Hogenboom. Using hierarchical edge bundles to visualize complex ontologies in GLOW. In *ACM Symp. on Applied Computing (SAC)*, 2012.

[6] B. Jou, T. Chen, N. Pappas, M. Redi, M. Topkara, and S.-F. Chang. Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *ACM Intl Conf. on Multimedia (MM)*, 2015.

[7] S. Kriglstein and R. Motschnig-Pitrik. Knoocks: New visualization approach for ontologies. In *Intl Conf. on Info. Vis. (InfoVis)*, 2008.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Adv. in Neural Info. Processing Sys. (NIPS)*, 2012.

[9] S. Lohmann, S. Negru, F. Haag, and T. Ertl. Visualizing ontologies with VOWL. *Semantic Web*.

[10] E. Motta, P. Mulholland, S. Peroni, M. d'Aquin, J. M. Gomez-Perez, V. Mendez, and F. Zablith. A novel approach to visualizing and navigating ontologies. In *Intl Conf. on the Semantic Web (ISWC)*. Springer, 2011.

[11] N. Pappas, M. Redi, M. Topkara, B. Jou, H. Liu, T. Chen, and S.-F. Chang. Multilingual visual sentiment concept matching. In *ACM Intl Conf. on Multimedia Retrieval (ICMR)*, 2016.

[12] R. Plutchik. *Emotion: A Psychoevolutionary Synthesis*. Harper & Row, 1980.

[13] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journ. of Ameri. Soci. for Info. Sci. and Tech.*, 61(12), 2010.

[14] T. D. Wang and B. Parsia. *CropCircles: Topology Sensitive Visualization of OWL Class Hierarchies*. Springer, 2006.