In Search of Design Inspiration: A Semantic-Based Approach

Rossitza Setchi
School of Engineering,
Cardiff University,
Cardiff CF24 3AA, UK
e-mail: setchi@cf.ac.uk

Carole Bouchard
Laboratory of New Product Design and Innovation,
Ecole Nationale Supérieure des Arts et Métiers,
75013 Paris, France
e-mail: carole.bouchard@paris.ensam.fr

Sources of inspiration help designers to define the context of their designs and reflect on the emotional impact of their new products. By observing and interpreting sources of inspiration, designers form vocabularies of terms, pallets of colors, or mood boards with images, which express their feelings, inspire their creativity and help them communicate design concepts. These ideas are the motivation behind the EU-funded project TRENDS, which aimed at developing a software tool that supports the inspirational stage of design by providing designers of concept cars with various sources of inspiration. This paper concentrates on OntoTag, the semantic-based image retrieval algorithm developed within the TRENDS project, and its evaluation. OntoTag uses concepts from a general-purpose lexical ontology called OntoRo, and semantic adjectives from a domain-specific ontology for designers called CTA, to index the images in the TRENDS database in a way which provides designers with a degree of serendipity and stimulates their creativity. The semantic-based algorithm involves the following four steps: (i) creating a collection of documents and images retrieved from the web, (ii) for each document, identifying the most frequently used keywords and phrases in the text around the image, (iii) identifying the most powerful concepts represented in each document, and (iv) ranking the concepts identified and linking them to the images in the collection. OntoTag differs significantly from earlier approaches as it does not rely on machine learning and the availability of tagged corpuses. Its main innovation is in the use of the words' monosemy and polysemy as a measure of their probability to belong to a certain concept. The proposed approach is illustrated with examples based on the software tool developed for the needs of two of the industrial collaborators involved in the TRENDS project. [DOI: 10.1115/1.3482061]

Keywords: creativity, design, inspiration, information retrieval, concept indexing, semantics, ontology

1 Introduction

The role of sectors of influence and sources of inspiration in creative design was understood relatively recently when research in this area focused on inspiration as a means of stimulating originality and creativity. The term *source of inspiration* refers to all conscious uses of previous designs and other objects and images in a design process [1]. It was found that sources of inspiration help designers to define the context of new designs, inform their creation, and reflect on their emotional impact. By observing and interpreting sources of inspiration, designers form vocabularies of terms, pallets of colors, or mood boards with images, which express their feelings, inspire their creativity, and help them communicate ideas to colleagues and clients.

This paper contributes to the research in this field by proposing a semantic indexing method developed in the context of a large scale collaborative research project called TRENDS. The project involved partners from four European countries that specialized in automotive design, content-based retrieval of images, search engines, intelligent agents, semantic-based systems, human-computer interaction, and software design. The aim of the project was to develop a software tool, which supports the inspirational stage of design. Developed for the needs of designers of concept cars, the tool assists them in collecting sources of inspiration from 12 sectors of influence.

The specific challenge addressed by the TRENDS project was the realization that "what is normally considered to be an effective information retrieval algorithm (i.e., producing high precision and recall) may be rather poor for serendipity, and thereby creativity" [2]. As reported in the same paper, when the participants in the studies were asked to use computer search facilities to locate inspirational images for a car design task, they sometimes used one car-related keyword to "anchor" their search to the design domain but coupled this with more semantically distant search terms (e.g., "car"+"exciting"+"futuristic"; "bold"+"colourful"+"cars"). This observation is indirectly supported by Ansburg and Hill [3] who stated that creative thinkers take advantage of incidentally presented cues, whereas analytical thinkers would not because solutions to analytical problems require focus on the problem elements. This statement is also consistent with Ford [4] who stated that "the level of creativity is dependent on the level of abstraction of the entities concerned."

This paper concentrates on the core innovation of the TRENDS computer support tool, which is its semantic-based image retrieval algorithm, which uses inspirational concepts and semantic adjectives and provides designers with a degree of diversity, ambiguity, and an opportunity to maintain an element of uncertainty in the search process.

The paper is organized as follows. Section 2 presents the state-of-the-art in two areas related to the problem domain of the paper. It first highlights the inspirational stage and the information requirements of the designers involved in conceptual design. It then reviews two areas closely associated with the semantic-based approach proposed in this research, namely, (i) content-based retrieval of images and (ii) semantic indexing and ontology-based annotation. Section 3 describes the semantic indexing method developed and its evaluation. The proposed approach is illustrated with examples using the TRENDS software developed for the needs of Fiat, Italy and Stile Bertone, Italy, which are two of the industrial collaborators involved in the TRENDS project. Section 4 presents conclusions and directions for future research.

Table 1 Sectors of influence identified by concept designers

Year	1997	2006
Designers	40 (10 professionals, 30 students)	30 professionals
Nationality	French, English, German	Italian, German, British, French
Sectors	1. Car design and automotive 2. Aircrafts, aeronautics 3. Architecture 4. Interior design and furniture 5. Hi-Fi 6. Product design 7. Fashion 8. Animals 9. Plants 10. Science Fiction 11. Virtual reality 12. Fine arts 13. Cinema 14. Music 15. Travels 16. Food	1. Car design and automotive 2. Architecture 3. Interior design and furniture 4. Fashion 5. Boat 6. Aircraft 7. Sport goods 8. Product design 9. Cinema and commercials 10. Nature and urban ambiances 11. Transportation 12. Music 13. Fine arts 14. Luxury brands 15. Animals 16. Packaging and advertising
Reference	Bouchard [6]	Mougenot et al. [7]

2 State-of-the-Art Review

The state-of-the-art review focuses first on the inspirational stage of design and its information requirements. It then reviews two semantic-based approaches, namely, semantic-based retrieval of images and ontology-based indexing and annotation.

2.1 Design Inspiration. Due to the open-mindedness needed in creative design, it is often considered positioned between arts and engineering. Similar to the arts, design requires extensive use of inspirational sources normally not supplied with the design brief.

The inspirational process is a crucial part of the design activity. Indeed, the novelty of each design candidate is primarily dependent on the inspirational process and on the way information is integrated during the generation of new solutions. However, even though the availability of inspirational materials (e.g., images, textures) is crucial, only a few research studies so far in the discipline of design science have been specifically centered on information collection in support of the inspirational phase of design. Nevertheless, this is an area of increasing importance and a number of researchers actively contribute to the field. A study comparing creative and analytical thinkers shows that creative thinkers usually employ data that is not directly linked to the problem to be solved [3]. This is why sources of inspiration, references, and influences play such a major role in creative design, both in defining the context for new designs and in informing their creation [5]. Designers use a large variety of artifacts and information coming from various areas. This includes precedents, artifacts, and images of art, human beings, natural phenomena, and everyday life. The design practices of concept cars designers were thoroughly investigated in previous studies conducted in 1997 and 2006 [6,7]. Both studies were based on interviews and observations during sketching activities. The studies showed that designers frequently use a number of sectors of influence. These are summarized in Table 1. The studies confirmed that these sectors can be considered of relevance when structuring an inspirational design database.

To finalize the inspiration phase, designers structure their inspiration sources by building trend or mood boards in various favorable contexts. Trend boards are collections of images compiled with the intention of provoking and communicating a trend or ambience during the product design process. The materials used in the elaboration of the trend boards are extracted from the sectors of influence (i.e., arts, nature, architecture, etc.). Studies show that trend boards offer a visual and sensorial channel of inspiration and communication [8] as they are more natural and empathic than verb-centric only approaches [9].

In the design process, sources of inspiration are essential for analogical [10,11] as well as divergent thinking [4] as they enable designers to build mental representations of both the problem space and the candidate solutions. The process of idea association by applying similarity, contrast, and contiguity principles is also supported by sources of inspiration [12].

Indeed, particular ideas exist in all sectors of influence. Examples include ideas expressing perceptions (e.g., modernity, freshness, impressions, and feelings), styling elements (aesthetics, proportion, structure, shape, line, and volume), technical characteristics (aerodynamics, engineering performance, technologies, capabilities, and materials), and consumer-related features (values, ergonomics, semantics, and sensory characteristics). Designers often project their own individuality through the object they create. This phenomenon contributes to originality and the creation of unique individual styles. For example, certain concepts are endowed with a signature by which the creator of the form can be recognized.

Sources of inspiration are collected through a deliberate information search effort carried out in order to meet the specifications of the brief combined with more or less conscious reactivation of knowledge previously memorized during various activities, possibly beyond the professional context [13]. The result of the information search is normally translated using visual means (trend panels, various sketches, etc.), text being used to highlight key points. At this stage, designers perform more or less systematic and organized activities ranging from simple browsing through collections to the creation of databases and trend panels. They often continue their studies in fields different from their original domain area (e.g., designers of concept cars considering trends in the design of boats). This process is highly subjective and personal. It is further supported by visits to showrooms and exhibitions as well as activities in fields of personal interest which designers refer to less regularly [14].

When collecting inspirational materials, designers use semantic adjectives in order to link words and images [2,15]. This activity is very specific and is normally supported by image and text search engines.

Designers recognize that their work deals with emotional content, although the process is not necessarily explicit. The pictures they select when they deliberately or unconsciously search for inspirational sources often have a high emotional impact [16]. In the discipline of design science, semantic, cognitive, and affective elements are considered closely related and constitute a significant field of current research. However, research centered on information retrieval has not considered until now the semantic dimensions of the inspirational process.

A recent study [17] suggests that the usefulness of the information captured in a design history depends on its indexing. Designers need help in information gathering, where they have to manage and categorize a huge amount of data. A study [18] confirms that compared with indexing based on low-level image features, indexing using keywords improves image retrieval as it is much more suitable in terms of human similarity perception. A work in that direction is reported in Ref. [19], which combines words referring to specific objects with words that express ambiguous Kansei feelings. However, as already mentioned, the majority of the image retrieval information systems currently used in the industry only employ low-level image features, whereas designers may wish to retrieve images using abstract concepts. With the trend emerging in the automotive industry to integrate Kansei engineering with AI and knowledge-based systems [20], it is apparent that future content-based image retrieval systems should move toward providing semantic-based image retrieval functions.

2.2 Semantic-Based Retrieval of Images. Research in content-based image retrieval is traditionally associated with image processing and low-level feature extraction. However, much of the research in image retrieval nowadays is focused on reducing the semantic gap between the low- level image features used in content-based retrieval and the high level concepts employed in queries [21,22]. As defined in Ref. [23], the semantic gap is, "… the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for the user in a given situation." In other words, the semantic gap is the discrepancy between the limited descriptive power of the low-level image features and the richness of the user's semantics.

Most approaches in image retrieval are based on keywords that either correspond to identifiable items describing the visual content of an image or relate to the context and the interpretation of that image. The user may be searching for "an image of" a particular entity (e.g. a flower or a child) or "an image about something" (e.g., an adventurous trip abroad). Advances in image analysis, object detection, and classification techniques may facilitate the automatic extraction of the first type of keywords. However, keywords belonging to the second category are unlikely to be automatically obtained from images [24,25] because from the point of view of feature representation, there is no single visual feature, which best describes the image content [21]. A particular challenging aspect in this context is dealing with concepts, which have no visual appearance in the images. Examples include concepts related to categories such as time, space, events and their significance, as well as abstract terms and emotions [26]. Such concepts could be extracted from annotations or the text complementing the image.

Most advanced image retrieval approaches, including the one advocated by the TRENDS partnership, employ hybrid techniques, which combine the use of visual features and text [25].

2.3 Ontology-Based Indexing and Annotation. The automatic annotation and indexing of images has been recognized by many researchers as an important enabling technology for bridging the semantic gap [22]. Semantic technologies and metadata languages are used as they provide means for defining class terminologies with well defined semantics and flexible data models for representing metadata descriptions [27].

Semantic annotations aim to provide some formalization of the content of the documents as a prerequisite for more comprehensive management [28]. Semantic annotations can be added both to documents and portions of documents. The annotations normally used a controlled vocabulary and are linked to some semantic resources (dictionaries, lexicons, glossaries, or ontologies). For example, the SemTag algorithm [29] has been applied to a set of 264×10^6 pages generating 434×10^6 automatically disambiguated semantic tags. The lexicon used by SemTag contains 200,000 words. The accuracy of the information retrieval is 82.01%.

It must be noted, however, that different ontologies may not have the same degree of formality. Controlled vocabularies, dictionaries, thesauri, and taxonomies are some of the most lightweight ontology types that have been widely used in annotation. These forms of vocabularies are not strictly formal and the annotations produced using them are normally pointers to terms in the vocabulary, which can be used to improve the search by using synonyms, antonyms, hyponyms, and hypernyms [30]. On the other hand, heavyweight axiomatized and formal ontologies are employed to incorporate formal semantics in the description of documents' content [31]. However, most of the formal ontologies do not include the vast number of terms that a thesaurus has. As stated by Corcho [30], thesauri, controlled vocabularies, and heavyweight ontologies are complementary since the first two can be used to provide agreed terms in specific domains while the latter provides formal semantics and constraints evaluation.

Indexing large document and image collections using ontologies is a relatively new area of research. Ontological indexing is defined as linking the words in the indexed text to ontological concepts [32]. A second definition proposed by Setchi and Tang [33] suggests that words can be treated as entities or concepts. They define concept indexing as "the analytic process of identifying instances (*entities*) and abstract ideas (*concepts*) within a text document and linking them to ontological concepts." In this definition, an entity is an identifiable and discrete instance existing in a text document, while a concept is an abstract idea inferred or derived from specific instances. A concept index is a machine understandable index of entities and concepts contained in a document collection.

For example, the sentence "Automobile designers are looking beyond the baby-boomer generation for inspiration…" contains the following entities: "automobile," "designers," "baby-boomer," "generation," and "inspiration." At the same time, this sentence could be associated with abstract ideas such as "inspiration," "innovation," and "creativity," which are implied in the text. Note that "inspiration" could be used both as an entity (keyword) and a generic concept. Depending on their significance for the whole text, the concept index for this document may include all or some of these entities and concepts.

Concept indexing is useful both for representing a document using abstract terms and for assigning concepts to specific words in documents. The process normally involves the following two steps: (i) extracting entities from unstructured text-based content using a lexical ontology and (ii) identifying concepts with the help of a concept knowledge base. Once entities and concepts are isolated, they are used to build a concept index [34].

The semantic concepts could be extracted and identified by disambiguating the sense of each word using linguistic repositories [32,35], semantic repositories [36], or domain-specific ontologies [37]. As highlighted in Ref. [36], linguistic repositories such as *WordNet* [38] do not capture the semantic relationships between concepts. On the other hand, semantic repositories such as *OpenCyc* [39], developed to capture and represent common sense, do not contain linguistic relationships (e.g., whether two words are synonyms), and domain dependent repositories such as the *Gene* ontology [37] only represent certain aspects of a domain and not the complete domain.

Although it is difficult to compare the existing approaches in terms of their accuracy, it is worth noting the results achieved by the two concept indexing approaches capable of processing large scale document collections at word level. The first study [32] uses three ontologies (WordNet, OpenCyc, and *SUMO*) containing 4115 concepts to index texts word by word. The ontology mapping accuracy is 96.2% while the accuracy of the ontology tagging is estimated to be between 60% and 70%. Similar to this approach, the concept indexing algorithm developed by Setchi and Tang [33,34] supports part of speech tagging, word sense disambiguation, and indexing using concepts from a rich general-purpose ontology called *OntoRo*. The average accuracy reported ranges from 76.40% to 78.91%, with the highest accuracy achieved when 90% of the corpus is used for training.

Despite using the same lexical resource (OntoRo), the semantic indexing algorithm *OntoTag* described in this paper differs significantly from earlier approaches [33,34] as it does not rely on machine learning and the availability of tagged corpuses. Its main innovation is in the use of the words' monosemy and polysemy as a measure of their probability to belong to a certain concept.

2.4 Related Work. Two investigations were conducted at the beginning of the TRENDS project, namely, a survey of commercial systems used by designers and a review of relevant European projects. It was found that none of the many commercial systems reviewed supported ontology development. However, several European projects funded by Framework Programme 6 of the European Commission (VIKEF, WIDE, AIM@SHAPE and XMEDIA) [40–43] were working on integrating ontologies in industrial practices. The study concluded that even though these projects also worked on ontology development, semantic search, and annotation of images and text, none of them achieved to integrate the results of such technology and interface conception into a software application. Their scope was also not as wide as the scope of the TRENDS project that was working on 23 sectors; VIKEF was related to pertinent information for trade fairs, WIDE specialized in management of information, X-MEDIA was at a very early phase, and AIM-SHAPE was more oriented toward 3D models.

The two most ambitious projects among the four projects discussed, WIDE and VIKEF, were based on a semantic mind mapping of the results. The WIDE project investigated the application of semantic web technologies and methods in an "integrated, scalable and reconfigurable design information support management and knowledge sharing system dedicated to multidisciplinary design teams in the context of innovative product design activities" [42]. The system offers text and image annotation with elicitation, refinement, and reuse of ontologies, which support the human interpretation of information about the domain. The design ontology defines different types of knowledge used and generated during design and the relations between them. It is then used by the system to identify what kinds of knowledge were requested, modified, or shared by users and to identify what other kinds of design knowledge could be relevant. The meta level subsystem semantically processes queries and results using an internal domain ontology, thesauri, and dictionaries. The content level subsystem handles the storage and retrieval of content as well as the annotation metadata and the management of external information sources. WIDE was in use at Italdesign Guigaro. Although the WIDE software appeared useful, the lack of images made directly available to the designers was deemed a major drawback. The ontology built during the AIM@SHAPE was briefly considered to be used as a part of the TRENDS ontology. However, the AIM@SHAPE ontology is limited to the area of car body, while TRENDS investigates 22 sectors in addition to the automotive sector.

3 Semantic Indexing in TRENDS

3.1 The TRENDS Approach. The interviews conducted with the designers at the start of the TRENDS project [2] revealed that they define their searches using the following:

(i) design-specific elements such as shape, texture, and color
(ii) sectors of influence (e.g., "architecture," "nature," "toys" or "automotive")
(iii) keywords such as "boats" or "cars"
(iv) semantic adjectives, for instance, "fresh," "aggressive," "luxury," "comfort," or "soft"

The TRENDS software integrates all these search options, which have been developed collaboratively by all partners.

This section focuses on one element of the developed solution: The semantic indexing using the domain-specific adjectives employed in design. OntoTag, the algorithm developed to index the images in the TRENDS database, uses concepts from two ontologies, namely, a generic ontology called OntoRo [34] and a domain-specific ontology for designers called *CTA* [44]. The CTA ontology is a purpose-built ontology, which has been developed using the so called conjoint trend analysis (CTA) method defined by studying the cognitive activity of designers during the inspirational phase of design [45]. In this phase, designers integrate many categories of information that are gradually formalized as design solutions throughout the design process. This encompasses the information contained in the brief and emanates from sources of inspiration and data searched by the designers to complete the design brief. These information and data can be categorized into information obtained from other designers and the designer's own information and experience originated from their interaction with the surrounding world.

The CTA method is based on an exhaustive, rigorous, and externalized information search process, which is molded on the designers' natural cognitive processes. This method was developed in order to improve the design process by the formalization of trend boards and pallets. It takes place in the early phases of design and helps designers focus on colors, forms, and textures as shown in Fig. 1. The outputs are trend boards, which offer a good representation of the references used by the designers before idea generation.

New design trends are identified by studying the sectors of influence. Images and keywords from these sectors are used to compose ambiences. An ambience is a coherent composition of images and keywords expressing values, semantic descriptors, and images composed following harmony rules. It is normally built mentally by the designers. CTA helps to make this kind of representation explicit (see Fig. 1) by using low-level descriptions of design features such as color (hue, saturation, and brightness), shape, and texture as well as high-level descriptors such as semantic adjectives, metaphors, analogies to other domains, and affective responses encountered while visualizing low-level information [46]. High-level information can be more influential and inspirational and more relevant to the emotional requirements of the consumer of the product than more concrete and unambiguous information. The core skill of a designer is his/her ability to link low-level features such as shape, color, and texture with high-level features such as semantics or sociological values. The global and discrete design elements are then extracted from these ambiences under the form of pallets. The pallets put together specific attributes linked to particular data sets (e.g., common properties of images in a database) so that they can be used to inspire the design of new products.

The CTA approach is combined in the TRENDS project with additional features suggested by 30 professional designers from Italy, France, and the U.K. who interviewed at the start of the project [47]. The core functions and knowledge structures were formalized and verified through user-centered experiments. The aim was to analyze and describe the information gathering requirements of the designers with the purpose of producing a computer aided tool. The protocol included the following three tasks: (i) selecting inspirational images from magazines, (ii) free description of the images selected, and (iii) articulating the design features in these images.

With car design as a sector of reference, designers had to search several sectors of influence and select pictures they considered inspirational. The images selected were representative of the inspirational processes, which routinely take place in the design

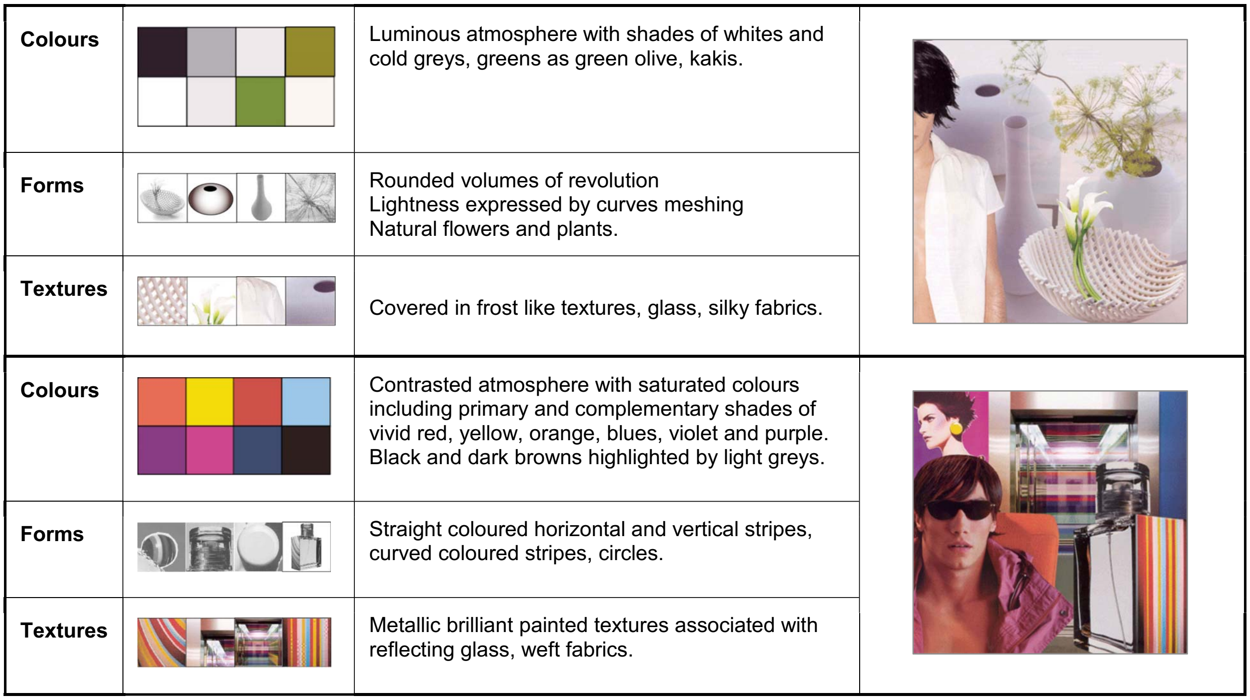


Fig. 1 Two trend boards created using conjoint trends analysis

process. Designers had to explain the reasons for choosing a particular image and to describe the images selected. The results were combined with specifications established from a lexical content analysis of previous CTA applications. The findings were exploited exhaustively when developing the database, the functionality, and the graphical user interface (GUI) and the algorithms of the TRENDS software. One of the results of this experiment directly related to the development of the semantic algorithm was capturing the specific relations between high and low-level information used by the designers interviewed (Table 1).

3.2 The TRENDS Ontologies. Nowadays, ontologies are the only widely accepted paradigm for the management of sharable and reusable knowledge in a way that allows its automatic interpretation [34]. They are collaboratively created across the web and used to index, search, and annotate documents. An ontology specifies a conceptualization of a domain in terms of concepts, attributes, and relations. Concepts are typically organized into a tree structure as well as linked through relations to form a semantic net structure.

3.2.1 The CTA Ontology for Designers. The CTA ontology is a purpose-built ontology, which has been developed using the CTA method described in the previous section. Fundamental to the CTA method is the establishment of a value-function-attributes chain, which uses semantic adjectives to link the marketing and design worlds. It has been found that the same semantic adjectives are used by designers when working with images and sketching new design concepts.

More or less consciously, designers follow a chaining process, which enables them to link consumer values and lifestyle information with low-level features, and *vice-versa*. This kind of chaining was made explicit in the field of advertising [48] and is regularly used in creating new messages and impressions by mixing words and images. Examples of such value-function-attributes chains in the context of the TRENDS study are the following sequences: "comfortable life" (end value)—"habitability" (function)—"roomy" (product attribute) and "security" (end value)—"assistance" (function)—"ergonomic" (product attribute).

The CTA ontology (Fig. 2) is defined in ontology web language (OWL) and developed in Protégé by creating instances and linking them using abstraction, aggregation, and dependency-based semantically rich relations.

As shown in Fig. 2, the top class "thing" contains five subclasses, namely, type, value, attribute, semantic adjective, and metaphor. In the figure, the asterisk sign " *" indicates that there may be zero or many instances for the corresponding attribute value. Class "value" defines the three types of values in CTA, which are "behavioral," "terminal," and "other." These three types are defined as three mutually exclusive instances of class type. "comfortableLifeValue" is an instance of class value. Class "attribute" has two subclasses, namely, "visualTactileAttribute" and "functionalAttribute." "colouredVis" is an instance of visualTactileAttribute and "durabilityFunc" is an instance of functionalAttribute. Class "semanticAdjective" is defined as a class whose instances are attributes in class value. "Ergonomic" is an instance of class semanticAdjective, which has several associated instances of semanticAdjective such as "soft," "light," "easy," etc. In particular, there is an attribute "hasWordId," which is linked to OntoRo described in the next section. The instances of class "metaphor" are also used as attribute values of value. The current version of the CTA ontology contains ten classes and 503 instances.

3.2.2 The Lexical Ontology OntoRo. OntoRo is a general-purpose lexical ontology based on Roget's thesaurus [49]. The Roget's thesaurus is a well known resource mainly used to facilitate the expression of ideas and assist in literacy composition. It is used in information retrieval to expand search items with other closely related words.

The decision to build and use OntoRo instead of employing WordNet or OpenCyc was primarily based on the fact that WordNet is a linguistic rather than a semantic repository and OpenCyc is a common sense ontology. To illustrate the differences, a simple experiment was conducted (Fig. 3), which compared the output of each of these ontologies. When an entry "attractive" is entered, OpenCyc links it to the concept "cuteness" and generates a list of words and phrases (e.g., "cute," "good looking," "cuter," "cutest,"

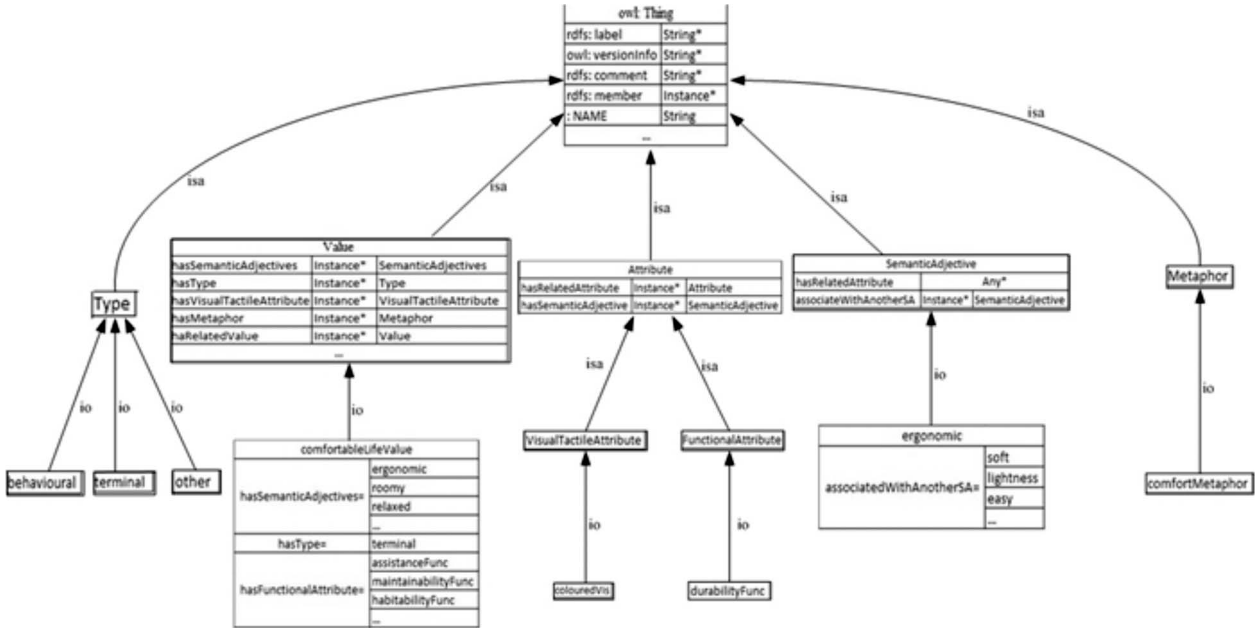


Fig. 2 CTA ontology

etc.), which are related to one contextual meaning of this word, which is as an attribute of human's personality. In addition to the meaning provided by OpenCyc, WordNet produces two more synsets for the word "attractive" and relates it to "having power to arouse interest" and "having the properties of a magnet." In addition, WordNet provides part of speech (POS) information and examples. OntoRo, on the other hand, links the word "attractive" to eight concepts (i.e., Nos. 178, 288, 291, 612, 826, 841, 859, and 887, see Fig. 3). One of these concepts, No. 841, belongs to class "emotion, religion, and morality" and section "personal emotion." It contains 451 semantically related words among which "beauty," "appeal," and "dress sense" as well as the less obvious associations "peacock," "Miss World," "Adonis," and "English rose." Concept No. 178 is an example of a completely different context. It belongs to class "abstract relations," section "causation" and contains 388 words such as "influence," "power," "magnetism," and "impact." Similar to WordNet, OntoRo also classifies the words using POS categories.

The two main arguments in support of the use of Roget's (and subsequently OntoRo) in the TRENDS application scenario are as follows. First, Roget's has a well established structure developed over more than 150 years, where the words/phrases are grouped and linked by their meaning. Second, although Roget's is organized hierarchically, it also maintains nonhierarchical, associative links between the words and phases in it.

These findings are consistent with the observations made by a number of researchers who have analyzed Roget's structure and semantic relations. For example, it is stated in Ref. [50] that, "in WordNet, only nouns are clearly organized in a hierarchy; verbs and adverbs are organized individually into various webs that are difficult to untangle." Next, as observed in Ref. [51], Roget's contains "semantic relations considered important for linguistic expression, which are not defined in other publicly available semantic networks, such as WordNet." This type of relation called "nonclassical" is further studied in Ref. [52], where it is emphasized that, "NLP methods and applications need to take account not only of "classical" lexical relations, as found in WordNet but the less structural, more context-dependent "nonclassical" relations that readers intuit in text." Furthermore, as they state, although Roget's is organized hierarchically, it also has a nonhierarchical, nonclassified "structure" for representing nonclassical relations such as "homeless"-"drunk," "rain"-"flood," and

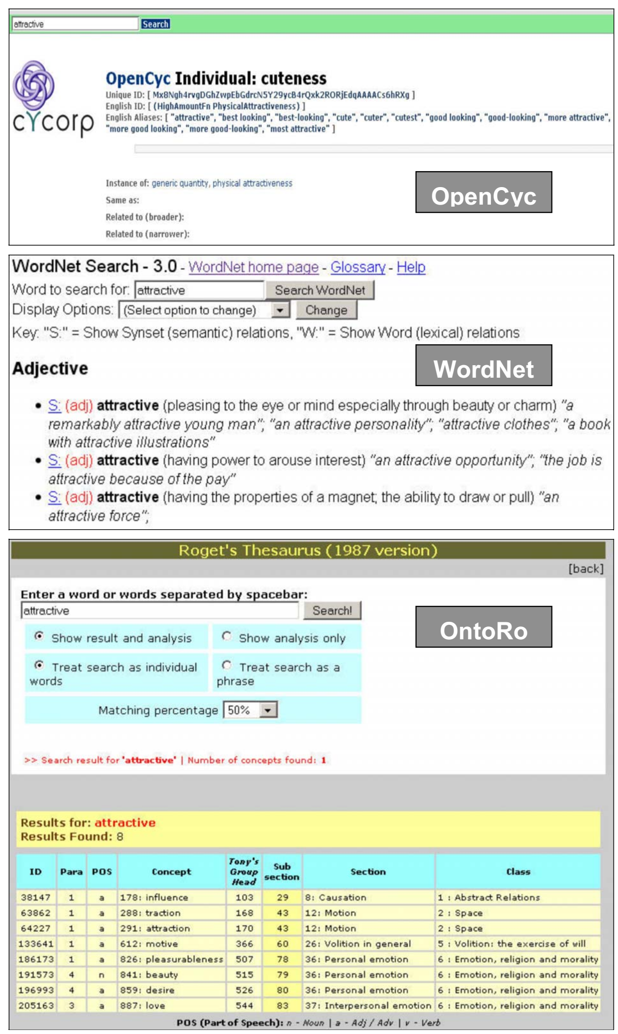


Fig. 3 OpenCyc, WordNet and OntoRo: a comparative example

"brutal"-"terrified."

The current version of OntoRo developed by Tang [53] includes 68,920 unique words and 228,130 entries, which are classified into 990 concepts, 610 head groups, 95 subsections, 39 sections, and six top level classes. OntoRo is also available as a web-based application [54].

3.3 Concept Indexing Algorithm. The concept indexing algorithm OntoTag is based on the assumption that there is a semantic link between the images in a web page and the text around them. The algorithm involves four steps.

Step (i): Retrieving web pages by targeted crawling and creating a collection of documents and images. This involves grabbing pages from websites and domains identified by the designers as their sectors of influence in a given context. For instance, the designers of concept cars listed 12 sectors of influence, among them "aerospace," "automobile," "architecture," "advertisement," "design," and "fashion."

Step (ii): Identifying the most frequently used keywords and phrases using the tf-idf function (1) [55].

$$w_{\text{tfidf}}(t_i,d_j) = \frac{\#(t_i,d_j)}{\#t_k} \cdot \log\frac{\#D}{\#(t_i,D)} \tag{1}$$

where $w_{\text{tfidf}}(t_i,d_j)$ (or tf-idf) represents the quantified weight of a term t_i contained in a document d_j from a collection D. This function embodies the idea that the more frequently a term occurs in a document and the fewer documents a term occurs in, the more representative it is of that document [55].

$\#(t_i,d_j)$ is the term count in the document d_j (i.e., the number of times a word appears in a document).

$\#\boldsymbol{t}_k$ is the number of occurrences of all terms in d_j (i.e., the total number of the terms in the document).

By dividing $\#(t_i,d_j)$ by $\#\boldsymbol{t}_k$, the term count is normalized to prevent a bias toward longer documents (which may have a higher term count regardless of the actual importance of that term in the document).

$$\frac{\#(t_i,d_j)}{\#t_k}$$

is the term frequency (often called tf), which is a measure of the importance of the term t_i within the document d_j.

$$\log\frac{\#D}{\#(t_i,D)}$$

the inverse document frequency (or idf) is a measure of the importance of the term in the collection D.

$\#D$ is the total number of documents in the collection, and

$\#(t_i,D)$ is the document frequency in the collection (i.e., the number of documents in the collection, which contain the term t_i).

Step (iii): Associating the most frequently used keywords and phrases with OntoRo and CTA concepts and computing the weight $w_{c_k}(d_j)$ of each concept C_k using Eq. (2).

$$w_{ck}(d_j) = \sum_{i=1}^{n}\left(k_{\text{CTA}} \cdot w_{\text{tfidf}}(t_i,d_j) \cdot \frac{1}{C_k(t_i)}\right) \tag{2}$$

where n is the number of terms t_i related to a concept C_k in a document d_j, $C_k(t_i)$ is the number of concepts C_k the term t_i is related to, $w_{\text{tfidf}}(t_i,d_j)$, computed using Eq. (1) in step (ii), denotes the significance of a certain term within a document, and k_{CTA} is a coefficient with two possible values, namely, 1.5 and 1. The first value (1.5) established using empirical studies is used if a certain term is contained in the CTA ontology to reflect its importance in the domain of interest. The value of k_{CTA} is 1 for all other terms.

The use of

$$\frac{1}{C_k(t_i)}$$

in Eq. (2) is based on empirical observations and the idea that monosemic words are more domain-oriented than polysemic ones and provides a greater amount of domain information. This converges with the common property of less frequent words being more informative, as they typically have fewer senses [56]. Therefore, the polysemy of each word can be used as a measure of the probability of the word to belong to a certain concept. Words that relate to one concept only are therefore more significant for that domain than words that relate to more concepts. This idea is further discussed in Sec. 3.4.

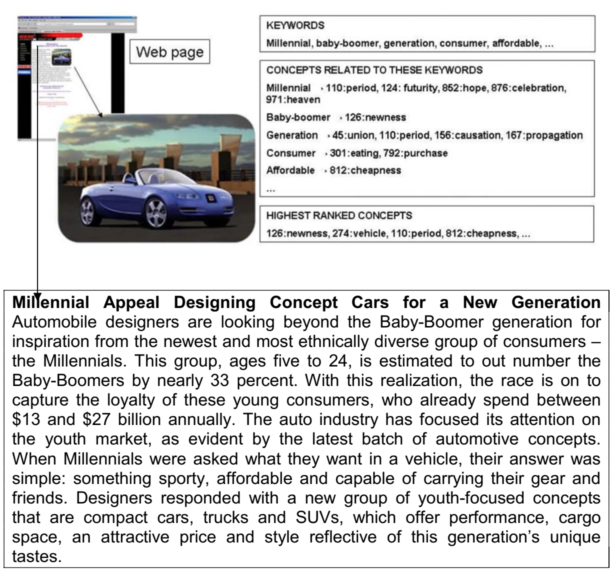
Step (iv): Ranking the concepts according to their significance $w_{ck}(d_j)$ and tagging the images with the concepts with the highest weight.

3.4 Computational Example. The operation of the concept indexing algorithm is illustrated in this section using a computational example based on a web page from the TRENDS collection (currently containing over 2×10^6 pages). The page shown in Fig. 4 contains the following text.

Millennial Appeal Designing Concept Cars for a New Generation *Automobile designers are looking beyond the baby-boomer generation for inspiration from the newest and most ethnically diverse group of consumers—the millennials. This group, ages 5–24, is estimated to outnumber the baby-boomers by nearly 33%. With this realization, the race is on to capture the loyalty of these young consumers, who already spend between $13 billion and $27 billion annually. The auto industry has focused its attention on the youth market, as evident by the latest batch of automotive concepts. When millennials were asked what they want in a vehicle, their answer was simple: something sporty, affordable and capable of carrying their gear and friends. Designers responded with a new group of youth-focused concepts that are compact cars, trucks and SUVs, which offer performance, cargo space, an attractive price and style reflective of this generation's unique tastes.*

The text is first preprocessed. Punctuation and numbers ("five," 24, and 33) are removed as well as all words with very high frequencies in the English language (such as "the," "of," "to," "a," "this," etc.). In addition, the words in plural (e.g., "designers"), gerund forms ("looking"), and past tense ("asked") are stemmed. The 148 words in the original text are thus reduced to the 85 words shown below.

Millennial Appeal Designing Concept Car New Generation Automobile designer look beyond baby-boomer generation inspiration new most ethnic diverse group consumer millennial group age estimate outnumber baby-boomer nearly percent realization race capture loyalty young consumer already spend between billion annual auto industry focus attention youth market evident latest batch automotive concept millennial ask want vehicle answer simple something sporty affordable capable carrying gear friend designer respond new group youth focus concept compact car truck SUV offer performance cargo space attractive price style reflective generation unique taste



Note: This picture and the corresponding text were published at http://ford.digitalsnippets.com/fiesta/. Since 2007 when the content was grabbed, Ford has updated their web page advertising Ford Fiesta so the page currently online has changed slightly. They still target the Millennials as a consumer group though.

Fig. 4 Example illustrating the ontology tagging algorithm

There are several words, which appear more than once in the text, namely, "baby-boomer," "concept," "consumer," "designer," "focus," "generation," "group," "millennial," "new," and "youth." The text box below contains the 69 unique words found in the page, arranged in alphabetical order.

Affordable age already annual answer appeal ask attention attractive auto automobile automotive baby-boomer batch between beyond billion capable capture cargo carrying compact concept car consumer designer designing diverse estimate ethnic evident focus friend gear generation group industry inspiration latest look loyalty market Millennial most nearly new offer outnumber percent performance price race realization reflective respond simple something space spend sporty style SUV taste truck unique vehicle want young youth

Table 2 shows the tf-idf calculated for each of these words. The tf value for any word occurring only once in the page is 1/85 =0.01176471, where 85 is the total number of the most representative words in the text. Words, which appear two and three times, have tf values 0.02352941 and 0.03529412, respectively. The idf values are computed on the basis of all documents in the collection (more than 2×10^6 documents). Frequently used terms such as "age," "group," "most," and "new" have relatively low idf values. The terms with the highest idf are "baby-boomer," "between," "beyond," "outnumber," "millennial," "reflective," and "sporty." Next, the tf-idf values for each word are calculated. The words with the highest tf-idf are "millennial" (0.310779348), "baby-boomer" (0.280387246), "generation" (0.153695318), "between" (0.124048822), "consumer" (0.123928701), "beyond" (0.117584063), "designer" (0.115958819), "outnumber" (0.099137732), etc. The first two words, "millennial" and "baby-boomer" have significantly higher tf-idf values compared with the remaining words in Table 2 due to the following main reasons, namely, their high idf values (i.e., relatively infrequent use in the collection) and high tf ("millennial" appears three times in the text, while "baby-boomer" occurs twice).

As Table 2 shows, each of the words in the text is related to one or more OntoRo concepts. In this example, all concepts are referred to using their concept number and the first word in the group of words related to that concept. For example, concept 126:newness contains 203 words and phrases listed in the box below.

newness, recently, recentness, recent date, recent occurrence, recent past, past time, present time, innovation, neoterism, neology, originality, novelty, gloss of novelty, freshness, dewiness, cleanness, greenness, immaturity, callowness, rawness, youth, renovation, restoration, renewal, resurrection, revival, clean slate, new leaf, new broom, modernism, modernity, modernness, modernization, up-to-dateness, topicality, contemporarily, present time, the latest, the latest thing, the in-thing, latest fashion, the last word, dernier cri, new look, contemporary style, trendiness, fashion, modernist, neologist, neologian, neoteric, neophiliac, futurist, advanced thinker, avant-garde, bright young thing, trendy, yuppie, baby-boomer, modern generation, younger generation, upstart, novus homo, parvenu, nouveau arrive, nouveau riche, vulgarian, Johnny-come-lately, incomer, new, newish, recent, of recent date, of recent occurrence, overnight, upstart, nouveau arrive, mushroom, novel, inventive, innovative, unprecedented, unheard of, original, brand-new, spick and span, like new, in mint condition, clean, green, evergreen, dewy, juicy, sappy, vernal, fresh, fresh as a daisy, fresh as paint, maiden, virgin, virginal, newborn, young, raw, unfledged, callow, immature, just out, just published, hot from the press, new-made, new-laid, straight from the oven, factory-fresh, untouched by human hand, unused, firsthand, untried, unbeaten, unexplored, unknown, untested, experimental, not broken in, not yet run in, budding, fledgling, aspiring, wannabe, beginning, modern, late, latter-day, contemporary, topical, present, up-to-the-minute, up-to-date, bang up-to-date, with it, a la mode, in the latest fashion, trendy, fashionable, ultramodern, modernistic, advanced, avant-garde, futuristic, untraditional, nontraditional, revolutionary, innovating, innovative, neoteric, newfangled, new-fashioned, state-of-the-art, cutting-edge, leading-edge, neological, modernized, renewed, renovated, rejuvenated, refurbished, repainted, restored, given a new look, given a new lease of life, brought up-to-date, revised, looking like new, freshened up, clean, modernise, do up, update, bring up-to-date, give a new lease of life, have the new look, go modern, go contemporary, get with it, move with the times, progress, newly, freshly, afresh, anew, like new, fresh-, new-, recently, overnight, just now, only yesterday, not long ago, a short time ago, lately, latterly, of late

Table 2 Linking semantic descriptors (high-level information) and design elements (low-level information)

Semantic descriptors	Design element
Simple, clean	Basic geometrical shapes, plain colors
Dynamic, active	Asymmetrical, tense lines
Relaxed, comfortable	Curves with big radius of curvature
Motionless	Symmetrical
Kitsch, loaded	Many objects
Free, unconventional	Nonregular forms/volumes
Quality, clean	Texture finishing with visual and tactile effects

Table 2 shows that some words from the page shown in Fig. 4 are related to one OntoRo concept only. For example, the word "affordable" has a single meaning related to concept 812:cheapness. Other monosemic words are "baby-boomer" (linked to 126:newness), "auto," "automobile," and "car" all related to 274:vehicle, "billion" to 99:number and "latest" to 121:present time. Most of the words in the text, however, are polysemic, with more than one meaning. Words with a very high number of related concepts include "answer" and "offer" (linked to 21 concepts), "compact" and "space" (related to 20 concepts), "estimate" and "inspiration" (19), and "new" (18). In addition, some of the words in the text appear in the CTA ontology of terms. These are "attractive," "inspiration," "new," "unique," and "youth."

Table 3 shows the next step of the semantic algorithm, where the weight of each concept is computed using Eq. (2). For example, the weight of the concept 812:cheapness is calculated on the basis that it is related to only one of the words in the text ("affordable"), which has a tf-idf value of 0.074895126 and does not belong to the CTA semantic adjectives (its K_{CTA} is 1.0).

$$\begin{aligned} w_{128:\text{cheapness}} &= k_{\mathrm{CTA}_{\text{affordable}}} \cdot w_{\mathrm{tfidf}_{\text{affordable}}} \cdot \frac{1}{C_{\text{affordable}}} \\ &= 1.0 \cdot 0.074895126 \cdot \frac{1}{1} = 0.074895126 \end{aligned}$$

The concept weight of 127:oldness, 131:age, 261:fold, 503:mental disorder, 650:health, 655:deterioration, and 842:ugliness is 0.002775065, which is the tf-idf value of the word " age," divided by the number of concepts "age" is related to (11).

$$\begin{aligned} w_{127:\text{oldness}} &= k_{\mathrm{CTA}_{\text{age}}} \cdot w_{\mathrm{tfidf}_{\text{age}}} \cdot \frac{1}{C_{\text{age}}} = 1.0 \cdot 0.030525712 \cdot \frac{1}{11} \\ &= 0.002775065 \end{aligned}$$

Some concepts are related to more than one word. For example, concept 445:appearance might be associated in some contexts with "look," "realization," and "space," while 821:excitation is linked to "appeal" and "inspiration." The word "inspiration" is an influential word in the domain of conceptual design and is part of the CTA ontology (its K_{CTA} coefficient is 1.5).

Table 3 An example showing the concepts related to the most representative words in the page shown in Fig. 4

t_i	$\#(t_i, d_j)$	$\#(t_i, d_j)/\#t_k$	$\log(\#D/\#(t_i, D)$	$(\#(t_i, d_j)/\Sigma t_k)^*$ $\log(\#D/\#(t_i, D)$	$C_k(t_i)$	Related concepts
Affordable	1	0.01176471	6.366085705	0.074895126	1	812:cheapness
Age	1	0.01176471	2.594685548	0.030525712	11	108:time, 110:period, 117:chronometry, 127:oldness, 131:age, 161:importance, 261:fold, 503:mental disorder, 650:health, 655:deterioration, 842:ugliness
Already	1	0.01176471	3.463151656	0.040742961	3	119:priority, 121:present time, 125:past time
Annual	1	0.01176471	3.414310733	0.040168362	6	110:period, 114:transience, 141:periodicity, 366:vegitable life, 528:publication, 589:book
Answer	1	0.01176471	4.591748279	0.054020568	21	24:agreement, 85:number, 148:reversion, 151:interchange, 156:causation, 157:effect, 460:answer, 475:reasoning, 479:computation, 520:interpretation, 584:interlocution, 588:correspondence, 623:plan, 635:sufficiency, 640:utility, 642:good policy, 658:remedy, 714:retaliation, 727:success, 839:wit, 878:insolence
Appeal	1	0.01176471	4.366118074	0.051366095	15	178:influence, 291:attraction, 532:affirmation, 533:negation, 583:allocution, 612:motive, 761:request, 762:deprecation, 781:giving, 804:payment, 821:excitation, 826:pleasurableness, 841:beauty, 887:love, 959:litigation
Ask	1	0.01176471	3.530525731	0.041535597	3	459:enquiry, 491:ignorance, 761:request
Attention	1	0.01176471	3.875599514	0.045595288	8	415:hearing, 447:intellect, 455:attention, 457:carefulness, 536:learning, 678:activity, 725:completion, 920:respect
Attractive	1	0.01176471	5.422452015	0.063793553	8	178:influence, 288:traction, 291:attraction, 612:motive, 826:pleasurableness, 841:beauty, 859:desire, 887:love
Auto	1	0.01176471	5.653271366	0.066509075	1	274:vehicle
Automobile	1	0.01176471	5.409926801	0.063646198	1	274:vehicle
Automotive	1	0.01176471	6.193054435	0.072859464	3	265:motion, 267:land travel, 274:vehicle
Baby-boomer	2	0.02352941	11.91645797	0.280387246	1	126:newness
Batch	1	0.01176471	6.294046241	0.074047603	2	26:quantity, 74:assemblage
Between	1	0.01176471	10.54414985	0.124048822	2	12:correlation, 231:interjacency
Beyond	1	0.01176471	9.994645369	0.117584063	3	34:superiority, 199:distance, 237:front
Billion	1	0.01176471	4.877966865	0.057387845	1	99:number
Capable	1	0.01176471	4.50745613	0.053028896	5	5:intrinsicality, 24:agreement, 160:power, 469:possibility, 498:intelligence
Capture	1	0.01176471	4.651339091	0.054721636	11	513:imagination, 551:representation, 590:description, 712:attack, 727:success, 729:trophy, 745:subjection, 747:restraint, 750:prisoner, 881:acquisition, 786:taking
cargo	1	0.01176471	5.358969635	0.063046702	5	32:greatness, 193:contents, 272:transference, 777:property, 795:merchandise
Carrying	1	0.01176471	4.570519443	0.053770817	4	167:propagation, 272:transference, 400:loudness, 404:resonance

Table 3 (***Continued.***)

t_i	$\#(t_i, d_j)$	$\#(t_i, d_j)/\#t_k$	$\log(\#D/\#(t_i, D)$	$(\#(t_i, d_j)/\Sigma t_k)^*$ $\log(\#D/\#(t_i, D)$	$C_k(t_i)$	Related concepts
Compact	1	0.01176471	5.483964561	0.06451723	20	24:agreement, 33:smallness, 48:coherence, 62:arrangement, 78:inclusion, 88:unity, 194:receptacle, 196:littleness, 198:contraction, 204:shortness, 324:density, 466:evidence, 488:assent, 569:conciseness, 764:promise, 765:compact, 767:security, 770:compromise, 791:barter, 843:beautification
Concept	2	0.02352941	3.832752199	0.090182405	3	451:idea, 485:belief, 513:imagination
Car	1	0.01176471	3.618070526	0.042565536	1	274:vehicle
Consumer	2	0.02352941	5.266969801	0.123928701	2	301:eating, 792:purchase
Designer	2	0.02352941	4.928249807	0.115958819	7	164:production, 192:abode, 228:dressing, 243:form, 556:artist, 594:drama, 623:plan
Designing	1	0.01176471	5.89705405	0.069377106	4	243:form, 541:falsehood, 930:improbity, 932:selfishness
Diverse	1	0.01176471	4.800244213	0.056473461	4	15:difference, 17:nonuniformity, 19:dissimilarity, 82:multiplicity
Estimate	1	0.01176471	5.367284209	0.06314452	19	62:arrangement, 86:numeration, 447:intellect, 449:thought, 463:discrimination, 465:measurement, 480:judgement, 507:expectation, 520:interpretation, 524:information, 532:affirmation, 557:language, 591:dissertation, 617:intention, 691:advice, 808:accounts, 809:price, 858:caution, 923:approbation
Ethnic	1	0.01176471	4.601831161	0.05413919	4	11:consanguinity, 169:parentage, 191:dweller, 371:humankind
Evident	1	0.01176471	5.532669839	0.065090233	4	443:visibility, 473:certainty, 478:demonstration, 522:manifestation
Focus	2	0.02352941	3.751268108	0.088265132	15	5:intrinsicality, 24:agreement, 45:union, 74:assemblage, 76:focus, 184:spatial, 192:abode, 225:centrality, 293:cenvergence, 438:sight, 617:intention, 708:party, 724:arena, 796:market, 882:sociality
Friend	1	0.01176471	3.709237901	0.043638093	9	89:accompaniment, 707:auxiliary, 880:friendship, 882:sociality, 897:benevolence, 903:benefactor, 965:divineness, 976:Orthodoxy, 978:sectarianism
Gear	1	0.01176471	5.526310517	0.065015418	8	74:assemlage, 77:classification, 194:receptacle, 228:dressing, 243:form, 630:tool, 777:property, 844:ornamentation
Generation	3	0.03529412	4.354700667	0.153695318	4	45:union, 110:period, 156:causation, 167:propagation
Group	3	0.03529412	2.194135972	0.077440093	16	50:combination, 53:part, 62:arrangement, 74:assemblage, 77:classification, 80:specialty, 104:multitude, 123:synchronism, 222:crossing, 276:aircraft, 365:animality, 371:humankind, 554:sculpture, 708:party, 722:combatant, 978:sectarianism
Industry	1	0.01176471	3.326644416	0.039136993	4	164:production, 622:business, 678:activity, 682:exertion
Inspiration	1	0.01176471	5.035731359	0.059243898	19	156:causation, 178:influence, 352:air, 398:sound, 476:intuition, 498:intelligence, 513:imagination, 570:diffuseness, 609:spontaneity, 612:motive, 623:plan, 818:feeling, 821:excitation, 822:excitability, 949:drunkeness, 965:divineness, 975:revelation, 979:piety
Latest	1	0.01176471	4.690168151	0.055178449	1	121:present time
Look	1	0.01176471	3.358197477	0.048669529	10	18:similarity, 243:form, 438:sight, 445:appearance, 453:curiosity, 455:attention, 524:i nformation, 547:indication, 688:conduct, 858:caution

Table 3 (*Continued.*)

t_i	$\#(t_i, d_j)$	$\#(t_i, d_j)/\#t_k$	$\log(\#D/\#(t_i, D)$	$(\#(t_i, d_j)/\Sigma t_k)^*$ $\log(\#D/\#(t_i, D)$	$C_k(t_i)$	Related concepts
Loyalty	1	0.01176471	5.588223608	0.065743807	11	597:willingness, 721:submission, 739:obedience, 745:subjection, 768:observance, 887:love, 917:duty, 920:respect, 929:probity, 931:disinterestedness, 979:piety
Market	1	0.01176471	3.385909905	0.039834234	9	76:focus, 192:abode, 297:ingress, 522:manifestation, 791:barter, 792:purchase, 793:sale, 796:market, 797:money
Millennial	3	0.03529412	8.805414846	0.310779348	5	110:period, 124: futurity, 852:hope, 876:celebration, 971:heaven
Most	1	0.01176471	1.649410062	0.019404824	2	26:quantity, 32:greatness
Nearly	1	0.01176471	3.574187107	0.04204926	4	33:smallness, 52:whole, 200:nearness, 495:error
New	3	0.03529412	1.465603946	0.051727198	18	15:difference, 19:dissimilarity, 21:originality, 52:whole, 68:beginning, 84:nonformity, 113:long duration, 122:different time, 125:past time, 126:newness, 132:child, 135:earliness, 143:change, 149:revolution, 163:weakness, 491:ignorance, 611:desuetude, 674:nonuse
Offer	1	0.01176471	3.335299334	0.039238816	21	137:occasion, 289:approach, 299:reception, 512:supposition, 532:affirmation, 595:will, 605:choice, 612:motive, 633:provision, 671:attempt, 756:permission, 759:offer, 761:request, 764:promise, 766:conditions, 781:giving, 791:barter, 792:purchase, 813:liberality, 962:reward, 979:piety
Outnumber	1	0.01176471	8.426707187	0.099137732	2	104:multitude, 637:redundance
Performance	1	0.01176471	3.458637403	0.040689852	13	157:effect, 164:production, 412:music, 413:musician, 551:representation, 557:language, 594:drama, 676:action, 725:completion, 768:observance, 876:celebration, 917:duty, 988:ritual
Price	1	0.01176471	4.132285326	0.048615121	11	28:equality, 86:numeration, 87:list, 150:transform, 465:measurement, 644:goodness, 737:command, 808:accounts, 809:price, 811:dearness, 963:punishment
Race	1	0.01176471	3.411258637	0.040132455	14	11:consanguinity, 77:classification, 169:parentage, 170:posterity, 267:land travel, 277:velocity, 306:overstepping, 350:stream, 371:humankind, 680:haste, 688:conduct, 708:party, 716:contention, 837:amusement
Realization	1	0.01176471	6.642319455	0.078144935	10	1:existence, 154:event, 319:materiality, 445:appearance, 484:discovery, 488:assent, 490:knowledge, 725:completion, 771:acquisition, 818:feeling
Reflective	1	0.01176471	6.989606938	0.08223067	2	417:light, 449:thought
Respond	1	0.01176471	4.39688608	0.051728072	5	24:agreement, 460:answer, 706:cooperation, 710:concord, 818:feeling
Simple	1	0.01176471	4.03743988	0.047499293	15	33: smallness, 44: simpleness, 425: color, 487: credulity, 491: ignorance, 499: unintelligence, 516: intelligibility, 540: veracity, 573: plainness, 575: elegance, 658: remedy, 699: artlessness, 701: facility, 846: people of taste, 869: commonalty
Something	1	0.01176471	3.971985503	0.046729241	3	3:substantiality, 32:greatness, 319:materiality
Space	1	0.01176471	3.506751005	0.041255894	20	26: quantity, 32: greatness, 46: disunion, 62: arrangement, 73: term, 78: inclusion, 107: infinity, 108: time, 183: space, 184: spatial, 195: size, 199: distance, 201: interval, 263: opening, 340: air, 348: plain, 410: melody, 445: appearance, 587: print, 632: store
Spend	1	0.01176471	4.739311885	0.05575661	3	634:waste, 681:leisure, 806:expenditure

Table 3 (*Continued.*)

t_i	$\#(t_i, d_j)$	$\#(t_i, d_j)/\#t_k$	$\log(\#D/\#(t_i, D)$	$(\#(t_i, d_j)/\Sigma t_k)^*$ $\log(\#D/\#(t_i, D)$	$C_k(t_i)$	Related concepts
Sporty	1	0.01176471	8.102415369	0.095322534	2	837:amusement, 875:ostentation
Style	1	0.01176471	3.199239742	0.037638115	16	7: state, 18: similarity, 77: classification, 117: chronometry, 243: form, 366: vegetable life, 514: meaning, 555: engraving, 561: nomenclature, 566: style, 575: elegance, 579: speech, 624: way, 688: conduct, 841: beauty, 848: fashion
Taste	1	0.01176471	5.385985344	0.063364533	15	43: mixture, 299: reception, 301: eating, 374: sensibility, 376: pleasure, 386: taste, 390: savouriness, 459: enquiry, 461: empiricism, 463: discrimination, 575: elegance, 605: choice, 818: feeling, 824: pleasure, 859: desire
Truck	1	0.01176471	5.023872728	0.059104385	3	273:carrier, 274:vehicle, 791:barter
Unique	1	0.01176471	3.839424853	0.045169704	8	17:nonuniformity, 19:dissimilarity, 21:originality, 29:inequality, 80:specialty, 84:nonformity, 88:unity, 644:goodness
Vehicle	1	0.01176471	4.690168151	0.055178449	7	67:sequel, 192:abode, 267:land travel, 272:transference, 273:carrier, 274:vehicle, 628:instrumentality
Want	1	0.01176471	3.912110957	0.046024835	11	35: inferiority, 55: incompleteness, 307: shortfall, 596: necessity, 627: requirement, 636: insufficiency, 647: imperfection, 731: adversity, 761: request, 801: poverty, 859: desire
Young	1	0.01176471	2.941811799	0.034609551	12	120: posteriority, 126: newness, 128: morning, 130: youth, 132: child, 134: adultness, 162: strength, 163: weakness, 164: production, 170: posterity, 538: learner, 670: nonpreparation
Youth	2	0.02352941	3.878035613	0.091247897	7	68:beginning, 126:newness, 130:youth, 132:child, 134:adultness, 161:importance, 170:posterity

$$w_{821:\text{excitation}} = k_{\text{CTA}_{\text{appeal}}} \cdot w_{\text{tfidf}_{\text{appeal}}} \cdot \frac{1}{C_{\text{appeal}}} + k_{\text{CTA}_{\text{inspiration}}} \cdot w_{\text{tfidf}_{\text{inspiration}}} \cdot \frac{1}{C_{\text{inspiration}}}$$

$$= 1.0 \cdot 0.051366095 \cdot \frac{1}{15} + 1.5 \cdot 0.059243898 \cdot \frac{1}{19}$$

$$= 0.008101556$$

The final step of the OntoTag algorithm includes ranking the concepts according to their weight. In this particular example, the concepts with the highest weight are 126:newness (0.307135096) and 274:vehicle (0.224591393), and they are used to tag the image. This is a good result, which truly captures the meaning of the text. The first concept, 126:newness, is linked to four words of which one, "baby-boomer," is a monosemic word with a high tf-idf value and occurs twice in the page. The two other, "new" and "youth" appear more than once in the text and are regarded as CTA terms. The second concept, 274:vehicle, is linked to six words of which three have single conceptual meaning.

This example shows that the algorithm takes into consideration four main factors, namely, the number of occurrences of a term in the text, its significance for the whole collection, the number of concepts each term relates to, and its importance for the domain of interest.

This example also illustrates the added value of using concepts when tagging images. As already shown, the image in Fig. 4 has been tagged with concepts 126:newness and 274:vehicle. As a result, a semantic search using the term "fresh," one of the semantic adjectives often used by designers, which is related to concepts 15:different, 21:original, and 126:newness, would retrieve this image as relevant. A keyword-based search using the same term would naturally produce no result because the term is not contained in the text. Similarly, the use of the CTA adjectives "trendy," "yuppie," and "modern" in a keyword-based search would give no results. The same terms used in a semantic search would retrieve this page because these words are instances of 126:newness. This illustrates a process called semantic query expansion, where each query is also processed semantically using the same ontological resources.

Figure 5 shows the user interface of the TRENDS software and some of its functionality in support of the designer's creativity. The software facilitates search for images using content-based and semantic-based retrieval. It supports individual work (using personal spheres) and collaboration (through the use of collective spheres). The figure shows a semantic search using a combination of words ("fresh," "new," and "original"), which results in the retrieval of 1556 images of various sizes of which 213 are large images of high quality. Images selected from those retrieved from the software are retained in the spheres and can be viewed in a variety of modes including a slideshow. A particularly useful feature of the software is the color pallet generator, which produces a color pallet from a number of images suggested by the user. The colors and textures identified as corresponding to a specific theme or mood are further used to produce mood boards using image manipulation and editing techniques.

3.5 Evaluation. The evaluation of OntoTag involves several stages, which are described in detail in an extensive report available at the TRENDS website [57].

The first evaluation included an examination of the TRENDS collection and manual evaluation of 500 pages with the purpose of identifying those cases, which challenge the semantic indexing algorithms developed. The conclusions were used to improve the performance of the ontology tagger. The experiments showed the effort required to create a tagged corpus and problems in the manual evaluations mainly related to subjectivity and difficulties in defining what is considered relevant when looking for diverse information.

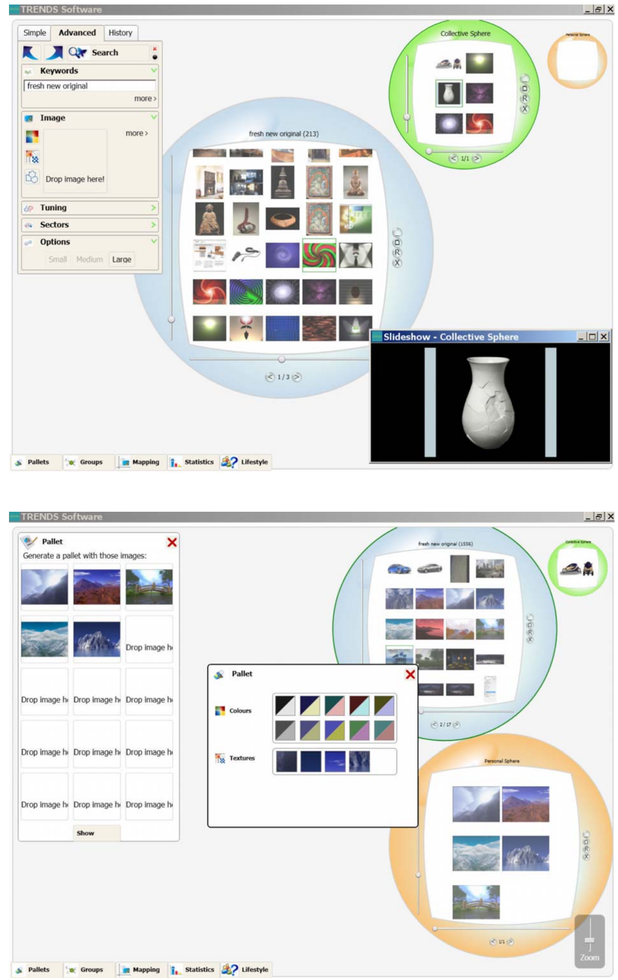


Fig. 5 Using semantic search to create a color pallet

The second evaluation was aimed at assessing the operational capability of the ontology tagger. A small collection of web pages was tagged in evaluation mode and checked manually to see if the corresponding output is tagged correctly from a semantic point of view. The results showed that seven out of ten pages were tagged with highly relevant concepts. For example, "9.html" (Fig. 6) is a web page describing a court lady in an oriental art collection, and this page is tagged with concepts 575:elegance, 837:amusement, 376:physical pleasure, 846:good taste, etc. Note that the word "elegantly" is mentioned once in this document ($\#(t_i, d_j) = 1$) and in the whole collection ($\#(t_i, D) = 1$), and this is the only document in the whole collection related to concept 846:good taste ($\#C_k(t_i) = 1$). Overall, the ontology tagger has produced a good result as it has captured the main concepts contained in the web page.

The aim of the next evaluation was to assess whether the information search improves when using concept or semantic adjective related queries by comparing OntoTag with a standard indexing desktop piece of software. A carefully composed collection of randomly selected pages from the sectors of influence was first built. The number of pages chosen per sector of influence was proportional to the number of pages in each of the sectors in the collection (i.e., aerospace is 28 pages, automobile is 27, architecture is 11, advertisement is ten, design is six, and fashion is two). To compare the performance of OntoTag, a keyword-based search engine had to be chosen. Three options were considered, namely, Google Desktop, Microsoft Windows Search and Copernic Desktop Search. *Copernic* [58] was chosen because (i) Google Desktop Search had a problem of excluding subdirectories for indexing

The standing figure of a female court attendant leans elegantly to one side, a small dog nestled into her voluminous sleeves. Fashionably attired in a high waisted gown of green and red-floret patterned fabric that falls well below the feet allowing only the exaggerated toes of her luxurious "cloud" slippers to peak out. The full green sleeves envelop her hands and arms which cradle a small, furry dog. The elaborate hairstyle, cold-painted black, of a high chignon, has been purposefully collapsed to emit an erotic recklessness. Her softly rounded face is modeled with expressive eyes under painted "moth" brows, a small, pert nose and vermillion lips.

#D:849

Words

word="elegantly"	Wtfidf(ti,dj)/howMany:2.928907690243953	Concept_id:846	#(ti,dj):1	#(ti,D):1	#Ck(ti):1
word="green"	Wtfidf(ti,dj)/howMany:0.10771405306221209	Concept_id:837	#(ti,dj):2	#(ti,D):49	#Ck(ti):22
word="well"	Wtfidf(ti,dj)/howMany:0.03313717946268596	Concept_id:642	#(ti,dj):1	#(ti,D):215	#Ck(ti):16
word="luxurious"	Wtfidf(ti,dj)/howMany:0.21356365678860703	Concept_id:376	#(ti,dj):1	#(ti,D):32	#Ck(ti):7
word="peak"	Wtfidf(ti,dj)/howMany:0.13483203316832856	Concept_id:725	#(ti,dj):1	#(ti,D):15	#Ck(ti):12
word="full"	Wtfidf(ti,dj)/howMany:0.037450095228675095	Concept_id:725	#(ti,dj):1	#(ti,D):196	#Ck(ti):16
word="green"	Wtfidf(ti,dj)/howMany:0.10771405306221209	Concept_id:837	#(ti,dj):2	#(ti,D):49	#Ck(ti):22
word="elaborate"	Wtfidf(ti,dj)/howMany:0.29289076902439526	Concept_id:725	#(ti,dj):1	#(ti,D):1	#Ck(ti):8
word="expressive"	Wtfidf(ti,dj)/howMany:0.4184153843205647	Concept_id:575	#(ti,dj):1	#(ti,D):1	#Ck(ti):7

Concepts

Concept No.:846 Wc(dj)=2.928907690243953
Concept No.:725 Wc(dj)=0.4651728974213989
Concept No.:575 Wc(dj)=0.4184153843205647
Concept No.:837 Wc(dj)=0.21542810612442417
Concept No.:376 Wc(dj)=0.21356365678860703
Concept No.:642 Wc(dj)=0.03313717946268596

Note. This web page is available at http://www.asianart.com/berwald/9.html

Fig. 6 OntoTag evaluation (example page 9.html, its concept tags, and their weighting, as generated by OntoTag)

and (ii) Microsoft Windows Search did not use all words in the web content to index the pages.

Using the same web document collection and given the same set of queries, the method evaluated how many relevant documents were returned by OntoTag and Copernic. Ten typical concepts from the CTA ontology were selected; they include abstract concepts used by designers (semantic adjectives) as well as abstract concepts relating to emotions. All results were manually examined by two experts. Table 4 shows the concepts used in the queries (e.g., "original," luxury," "passionate," etc.) and the results produced when using Copernic and OntoTag. *Precision* was used as a measure of relevance (defined as the fraction of all documents retrieved that are relevant to the user's query). The results in the table are shown as fractions; the numerator shows the number of all relevant documents retrieved and the denominator indicates the number of all documents retrieved.

The results in the table show the precision achieved by the Copernic and OntoTag algorithms. In seven cases out of the ten queries conducted, OntoTag gives equally good or better results. Some of the cases demonstrate high precision (see "high tech," "well-being," "luxury," etc.). Moreover, the analysis of the results clearly shows that the high precision of OntoTag is a result of the semantic approach used. For example, the search for the keyword "aggressive" (Fig. 7) does not produce any results with Copernic. In this case, OntoTag returns three applicable results out of all four results in total (the discarded one is considered "marginally" relevant). In this experiment, OntoTag considers the space shuttle, Porsche sport car, military personnel, and vehicles relevant to the concept "aggressive." Note that these web pages do not contain the words "aggressive," "aggression," etc. but the contextual paragraphs include statistically significant information to classify them as related to the concept "aggressive." Note that in the same example (Fig. 7), OntoTag produces four results but only three of them are considered relevant on the basis that the image of the flying helicopter does not invoke such a connotation. However, analysis of the OntoTag results show that although the picture may not be considered particularly of aggressive nature, the text on the page refers to military operations, which explains the tagging. On the other hand, although the text on the page containing the car is not explicitly related to any military themes, the expert considers the picture relevant to the query and indeed, an unexpected and rather encouraging result in the context of the TRENDS project.

However, as shown in Table 4, there are cases when Copernic retrieves some pages (even though with very low precision) while OntoTag returns a few or no results (see the query "original" in Table 4). This indicates that although semantic-based searches demonstrate clear advantages, they are not infallible and should therefore be used in combination with keyword-based search (Table 5).

In addition, OntoTag was evaluated through user-centered studies. Qualitative and quantitative methods were used to evaluate each of the three prototypes developed as well as the final software. The system was evaluated according to the main aspects identified by both the HCI experts and the focus groups as critical for the provision of a computer-based inspirational support. These relate to the quality of database content and the availability of high quality images, the performance of the search algorithms, the user-friendliness of the interface design, and the overall functionality of the system [59].

The evaluation protocol allowed the content of the database and the performance of the search algorithms to be assessed independently from the system interface. Twelve participants from the design departments of Fiat (6) and Stile Bertone (6) took part in these experiments. First, each designer was given a design brief and asked to write down the keywords they would use when searching for inspirational images. The design brief was as follows.

The owner of this car will be retired, married, countryside dweller and a dog owner. The customer's main requirement is that the car is economical and spacious. The car should be easy to get in and out, and easy to load. It should be functional and smart.

Next, the individuals conducting the experiment formed search queries using words suggested by the designers (i.e., "smart," "farm," "ecodesign," "comfort," etc.). Two different image search engines were used in this experiment, namely, TRENDS and Google image search. The designers were then presented with two sets of ten pages, each set retrieved using a different search engine. Each page contained 16 images as shown in Fig. 8. The designers were not told which set has been produced by which search engine. They were then asked to consider the design brief, to scroll through the images, and rate each image in terms of its inspirational value, usefulness, aesthetics, serendipity, closeness to the idea they have had in mind, and image quality. In addition, the designers were presented with the opportunity to conduct random search by browsing images in the collection. Figure 8 shows the evaluation sheet used by the designers during the tests and the evaluation results for the first criterion: inspirational value.

It is important to note that although the participants were asked

Table 4 Weight of the concepts related to the page shown in Fig. 4

C_k	t_i related to C_k	n related to C_k	W_{ck}
812:cheapness	Affordable	1	0.074895126
127:oldness, 131:age, 261:fold, 503:mental disorder, 650:health, 655:deterioration, 842:ugliness	Age	1	0.002775065
119:priority	Already	1	0.013580987
114:transience, 141:periodicity, 528:publication, 589:book	Annual	1	0.006694727
85:number, 148:reversion, 151:interchange, 475:reasoning, 479:computation, 584:interlocution, 588:correspondence, 635:sufficiency, 640:utility, 642:good policy, 714:retaliation, 839:wit, 878:insolence	Answer	1	0.002572408
533:negation, 583:allocution, 762:deprecation, 804:payment, 959:litigation	Appeal	1	0.051366095
415:hearing, 457:carefulness, 536:learning	Attention	1	0.005699411
288:traction	Attractive	1	0.011961291
265:motion	Automotive	1	0.024286488
12:correlation, 231:interjacency	Between	1	0.062024411
34:superiority, 237:front	Beyond	1	0.039194688
99:number	Billion	1	0.057387845
160:power, 469:possibility	Capable	1	0.010605779
590:description, 712:attack,729:trophy, 747:restraint, 750:prisoner, 881:acquisition, 786:taking	Capture	1	0.004974694
193:contents, 795:merchandise	Cargo	1	0.01260934
400:loudness, 404:resonance	Carrying	1	0.013442704
48:coherence, 196:littleness, 198:contraction, 204:shortness, 324:density, 466:evidence, 569:conciseness, 765:compact, 767:security, 770:compromise, 843:beautification	Compact	1	0.003225862
451:idea, 485:belief	Concept	1	0.030060802
556:artist	Designer	1	0.016565546
541:falsehood, 930:improbity, 932:selfishness	Designing	1	0.017344277
82:multiplicity	Diverse	1	0.014118365
480:judgement, 507:expectation, 591:dissertation, 691:advice, 923:approbation	Estimate	1	0.003323396
443:visibility, 473:certainty, 478:demonstration	Evident	1	0.016272558
225:centrality, 293:convergence, 724:arena	Focus	1	0.005884342
89:accompaniment, 707:auxiliary, 880:friendship, 897:benevolence, 903:benefactor, 976:Orthodoxy	Friend	1	0.004848677
630:tool, 844:ornamentation	Gear	1	0.008126927
50:combination, 53:part, 123:synchronism, 222:crossing, 276:aircraft, 554:sculpture, 722:combatant	Group	1	0.004840006
622:business, 682:exertion	Industry	1	0.009784248
352:air, 398:sound, 476:intuition, 570:diffuseness, 609:spontaneity, 822:excitability, 949:drunkeness, 975:revelation	Inspiration	1	0.004866953
18:similarity, 453:curiosity, 547:indication	Look	1	0.004866953
597:willingness, 721:submission, 739:obedience, 929:probity, 931:disinterestedness	Loyalty	1	0.00597671
297:ingress, 793:sale, 797:money	Market	1	0.004426026

Table 4 (*Continued.*)

C_k	t_i related to C_k	n related to C_k	W_{ck}
124:futurity, 852:hope, 971:heaven	Millennial	1	0.06215587
200:nearness, 495:error	Nearly	1	0.010512315
113:long duration, 122:different time, 135:earliness, 143:change, 149:revolution, 611:desuetude, 674:nonuse	New	1	0.0043106
137:occasion, 289:approach, 512:supposition, 595:will, 633:provision, 671:attempt, 756:permission, 759:offer, 766:conditions,813:liberality, 962:reward	Offer	1	0.001868515
637:redundance	Outnumber	1	0.049568866
412:music, 413:musician, 676:action, 988:ritual	Performance	1	0.003129989
28:equality, 87:list, 150:transform, 737:command, 811:dearness, 963:punishment	Price	1	0.004419556
11:consanguinity, 169:parentage, 277:velocity, 306:overstepping, 350:stream, 680:haste, 716:contention	Race	1	0.002866604
1:existence, 154:event, 484:discovery, 490:knowledge, 771:acquisition	Realization	1	0.007814493
417:light	Reflective	1	0.041115335
706:cooperation, 710:concord	Respond	1	0.010345614
44: simpleness, 425: color, 487:credulity, 499: unintelligence, 516: intelligibility, 540: veracity, 573: plainness, 699: artlessness, 701: facility, 846: people of taste, 869: commonalty	Simple	1	0.004749929
3:substantiality	Something	1	0.015576414
46:disunion, 73:space, 107:infinity, 183:space, 195:size, 201:interval, 263: opening, 340: air, 348: plain, 410: melody, 587: print, 632: store	Space	1	0.002062795
634:waste, 681:leisure, 806:expenditure	Spend	1	0.018585537
875:ostentation	Sporty	1	0.047661267
7: state, 18: similarity, 514: meaning, 555: engraving, 561:nomenclature, 566: style, 579: speech, 624: way, 848: fashion	Style	1	0.003528573
43: mixture, 374: sensibility, 376: pleasure, 386: taste, 390: savouriness, 461: empiricism, 824: pleasure	Taste	1	0.006336453
29:inequality	Unique	1	0.029552193
67:sequel, 628:instrumentality	Vehicle	1	0.005646213
35: inferiority, 55: incompleteness, 307: shortfall, 596: necessity, 627: requirement, 636: insufficiency, 647: imperfection, 731: adversity, 801: poverty	Want	1	0.007882636
120: posteriority, 128: morning, 162: strength, 538: learner, 670: nonpreparation	Young	1	0.004184076
Two terms			
117:chronometry	Age, style	2	0.006303638
161:importance	Age, youth	2	0.022328185
108:time	Age, space	2	0.004837859
121:present time	Already, latest	2	0.068759436
125:past time	Already, new	2	0.017891587
366:vegetable life	Annual, style	2	0.0102233
157:effect	Answer, performance	2	0.005702397
460:answer	Answer, respond	2	0.012918022
520:interpretation	Answer, estimate	2	0.005895804

Table 4 (*Continued.*)

C_k	t_i related to C_k	n related to C_k	W_{ck}
658:remedy	Answer, simple	2	0.007322337
727:success	Answer, capture	2	0.007547102
291:attraction, 826:pleasurableness	Appeal, attractive	2	0.015385698
781:giving	Appeal, offer	2	0.005292921
821:excitation	Appeal, inspiration	2	0.008101556
459:enquiry	Ask, taste	2	0.020181652
447:intellect	Attention, estimate	2	0.009022807
678:activity	Attention, industry	2	0.015483659
445:attention	Attention, look	2	0.010566364
920:respect	Attention, loyalty	2	0.011676121
199:distance	Beyond, space	2	0.041257482
5:intrinsicality	Capable, focus	2	0.016490121
498:intelligence	Capable, inspiration	2	0.015282929
551:representation	Capture, performance	2	0.008104683
745:subjection	Capture, loyalty	2	0.010951404
777:property	Cargo, gear	2	0.020736268
167:propagation	Cargo, generation	2	0.05103317
78:inclusion	Compact, space	2	0.005288656
88:unity	Compact, unique	2	0.011695118
194:receptacle	Compact, gear	2	0.011352789
488:assent	Compact, realization	2	0.011040355
764:promise	Compact, offer	2	0.005094377
301:eating	Consumer, taste	2	0.068300804
228:dressing	Designer, gear	2	0.024692473
594:drama	Designer, performance	2	0.019695534
15:difference	Diverse, new	2	0.018428965
17:nonuniformity	Diverse, unique	2	0.022587685
524:information, 858:caution	Estimate, look	2	0.008190349
86:numeration, 465:measurement, 808:accounts, 809:price	Estimate, price	2	0.007742952
449:thought	Estimate, reflective	2	0.044438731
463:discrimination	Estimate, taste	2	0.009659849
557:language	Estimate, Performance	2	0.006453384
617:intention	estimate, focus	2	0.009207738
522:manifestation	Evident, market	2	0.020698584
45:union	Focus, generation	2	0.044308172
438:sight	Focus, look	2	0.010751295

Table 4 (*Continued.*)

C_k	t_i related to C_k	n related to C_k	W_{ck}
76:focus, 796:market	Focus, market	2	0.010310368
184:spatial	Focus, space	2	0.007947137
882:sociality	Focus, friend	2	0.010733019
965:divineness	Friend, inspiration	2	0.009525827
978:sectarianism	Friend, group	2	0.009688683
80:specialty	Group, unique	2	0.013309325
104:multitude	Group, outnumber	2	0.054408872
371:humankind	Group, race	2	0.00770661
18:similarity	Look, style	2	0.008395526
768:observance, 917:duty	Loyalty, performance	2	0.009106698
821:excitation	Appeal, inspiration	2	0.008101556
459:enquiry	Ask, taste	2	0.020181652
447:intellect	Attention, estimate	2	0.009022807
678:activity	Attention, industry	2	0.015483659
445:attention	Attention, look	2	0.010566364
920:respect	Attention, loyalty	2	0.011676121
199:distance	Beyond, space	2	0.041257482
5:intrinsicality	Capable, focus	2	0.016490121
498:intelligence	Capable, inspiration	2	0.015282929
551:representation	Capture, performance	2	0.008104683
745:subjection	Capture, loyalty	2	0.010951404
777:property	Cargo, gear	2	0.020736268
167:propagation	Cargo, generation	2	0.05103317
78:inclusion	Compact, space	2	0.005288656
88:unity	Compact, unique	2	0.011695118
194:receptacle	Compact, gear	2	0.011352789
488:assent	Compact, realization	2	0.011040355
764:promise	Compact, offer	2	0.005094377
301:eating	Consumer, taste	2	0.068300804
228:dressing	Designer, gear	2	0.024692473
594:drama	Designer, performance	2	0.019695534
15:difference	Diverse, new	2	0.018428965
17:nonuniformity	Diverse, unique	2	0.022587685
524:information, 858:caution	Estimate, look	2	0.008190349
86:numeration, 465:measurement, 808:accounts, 809:price	Estimate, price	2	0.007742952
449:thought	Estimate, reflective	2	0.044438731

Table 4 (*Continued.*)

C_k	t_i related to C_k	n related to C_k	W_{ck}
463:discrimination	Estimate, taste	2	0.009659849
557:language	Estimate, performance	2	0.006453384
617:intention	Estimate, focus	2	0.009207738
522:manifestation	Evident, market	2	0.020698584
45:union	Focus, generation	2	0.044308172
438:sight	Focus, look	2	0.010751295
76:focus, 796:market	Focus, market	2	0.010310368
184:spatial	Focus, space	2	0.007947137
882:sociality	Focus, friend	2	0.010733019
965:divineness	Friend, inspiration	2	0.009525827
978:sectarianism	Friend, group	2	0.009688683
80:specialty	Group, unique	2	0.013309325
104:multitude	Group, outnumber	2	0.054408872
371:humankind	Group, race	2	0.00770661
18:similarity	Look, style	2	0.008395526
768:observance, 917:duty	Loyalty, performance	2	0.009106698
876:celebration	Millennial, performance	2	0.065285858
52:whole	Nearly, new	2	0.014822915
21:originality, 84:nonformity	New, unique	2	0.012779919
68:beginning	New, youth	2	0.023863721
163:weakness	New, young	2	0.007194729
299:reception, 605:choice	Offer, taste	2	0.008204968
644:goodness	Price, unique	2	0.012888876
837:amusement	Race, sporty	2	0.050527871
319:materiality	Realization, something	2	0.023390907
273:carrier	Truck, vehicle	2	0.027584097
130:youth, 134:adultness	Young, youth	2	0.02243725
Three terms			
156:causation	Answer, generation, inspiration	3	0.045673387
623:plan	Answer, designer, inspiration	3	0.022256053
178:influence	Appeal, attractive, inspiration	3	0.020062847
532:affirmation	Appeal, estimate, offer	3	0.008616317
841:beauty	Appeal, attractive, style	3	0.018914271
887:love	Appeal, attractive, loyalty	3	0.021362407

Table 4 (*Continued.*)

C_k	t_i related to C_k	n related to C_k	W_{ck}
491:ignorance	Ask, new, simple	3	0.022905728
725:completion	Attention, performance, realization	3	0.016643893
859:desire	Attractive, taste, want	3	0.02248182
267:land travel	Automotive, race, vehicle	3	0.035035727
26:quantity	Batch, space, most	3	0.048789008
513:imagination	Capture, concept, inspiration	3	0.039712646
272:transferance	Cargo, carrying, vehicle	3	0.03393468
33:smallness	Compact, nearly, simple	3	0.018488106
792:purchase	Consumer, market, offer	3	0.068258892
19:dissimilarity	Diverse, new, unique	3	0.026898285
708:party	Focus, group, race	3	0.013590952
979:piety	Inspiration, loyalty, offer	3	0.012522375
132:child	New, young, youth	3	0.02674785
688:conduct	Look, race, style	3	0.01126213
445:appearance	Look, realization, space	3	0.014744241
170:posterity	Race, young, youth	3	0.025303854
575:elegance	Simple, style, taste	3	0.014614956
Four terms			
110:period	Age, annual, generation, millennial	4	0.110049491
612:motive	Appeal, attractive, inspiration, offer	4	0.021931362
761:request	Appeal, ask, offer, want	4	0.023322196
126:newness	Baby-boomer, new, young, youth	4	0.307135096
74:assemblage	Batch, focus, gear, group	4	0.055875077
32:greatness	Cargo, most, something, space	4	0.039950961
62:arrangement	Compact, estimate, group, space	4	0.013452058
791:barter	Compact, market, offer, truck	4	0.029221864
164:production	Designer, industry, performance, young	4	0.032363912
192:abode	Designer, focus, market, vehicle	4	0.034758549
77:classification	Gear, group, race, style	4	0.01936211
818:feeling	Inspiration, realization, respond, taste	4	0.029173711

Table 4 (***Continued.***)

C_k	t_i related to C_k	n related to C_k	W_{ck}
	Five terms		
24:agreement	Answer, capable, compact, focus, respond	5	0.032634005
243:form	Designer, designing, gear, look, style	5	0.050432276
	Six terms		
274:vehicle	Auto, automobile, automotive, concept car, truck, vehicle	6	0.224591393

to scroll through sets of ten pages, only the first six pages contained images selected as inspirational by all 12 participants. Figure 8 shows the mean for the inspirational value of the images included in page 1 and page 6. The TRENDS system seems to produce relatively more inspirational images in the first page and less image in the sixth page than Google. Overall, the TRENDS system was regarded as a useful tool.

Finally, tests with the final prototype were conducted with eight car designers, six from Fiat and two from Stile Bertone, who used the system as part of their normal working routine over a period of one month. They provided daily feedback on their experience when using the TRENDS system, in particular are the ease of use of its interface, the quality of the database content and search algorithms, and the usefulness of the system functions. These designers used a range of available functions including text-based and content-based retrieval. They raised a number of points for consideration such as the quality of the images, difficulty in finding some elements of the interface, and the slow time of response. The results of the final tests indicate the potential of the TRENDS software for further commercial exploitation and identify the outstanding issues which need to be addressed. End users considered the TRENDS functionality very useful in the design context. Overall, the semantic search was accepted well.

4 Conclusions and Future Work

The ontology tagger OntoTag developed within the TRENDS project has demonstrated good performance and scalability and has been integrated with keyword-based indexing and content retrieval algorithms in the final TRENDS prototype. After processing all words in the text, OntoTag extracts all related concepts and ranks them according to their significance. As a result, it produces a set of concept numbers for each text, which is then used to retrieve information in a process called semantic expansion, where a keyword query is also processed semantically. OntoTag differs significantly from earlier approaches as it does not rely on machine learning and the availability of tagged corpuses. Its main innovation is in the use of the words' monosemy and polysemy as a measure of their probability to belong to a certain concept.

This paper demonstrates that a concept-based search, combined with content-based image retrieval and keyword-based search, is a useful tool for designers involved in creative tasks. It complements traditional methods by providing images with a degree of diversity and high inspirational value. This paper highlights the potential of the semantic-based algorithms in applications supporting inspiration and creativity.

Further research is needed to study the impact of serendipity,

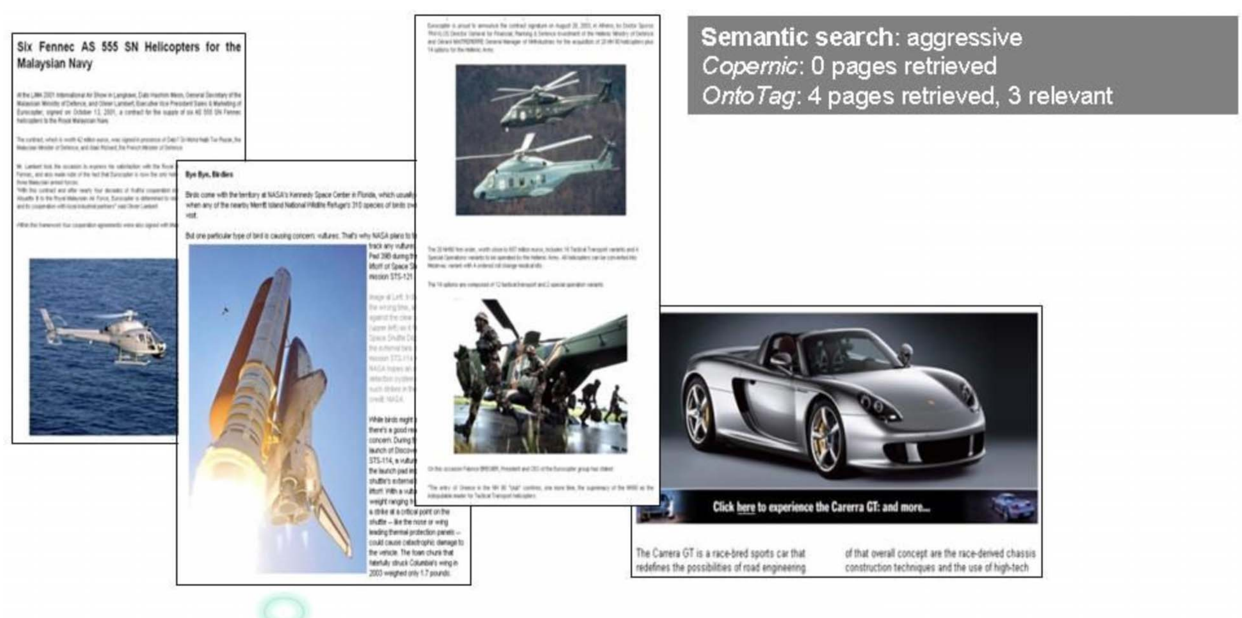


Fig. 7 OntoTag evaluation: semantic search for aggressive

Table 5 Evaluation results: OntoTag and Copernic compared

Concept Search Queries	PRECISION = $\frac{\text{number\_of\_relevant\_documents}}{\text{all\_documents\_retrieved}}$ Copernic Results	OntoTag Results
Adaptable	0/0	0/0
Aggressive	0/2	3/4
High tech	3/31	3/3
Original	1/12	0/0
Luxury	2/10	1/1
Passionate	0/0	0/0
Pleasant	1/1	3/3
Refreshing	0/0	2/2
Soft	3/4	1/1
Well-being	0/6	1/1

diversity, and semantic associations on imagination, inspiration, and creativity. An outstanding issue in the area of information retrieval is the lack of well established methodologies, indexed corpuses for comparative studies, and qualitative measures for evaluating semantic algorithms.

The concept indexing algorithm is currently being used as a research tool in three ongoing research projects, which explore its current limitations and the use of more complex ontological relationships.

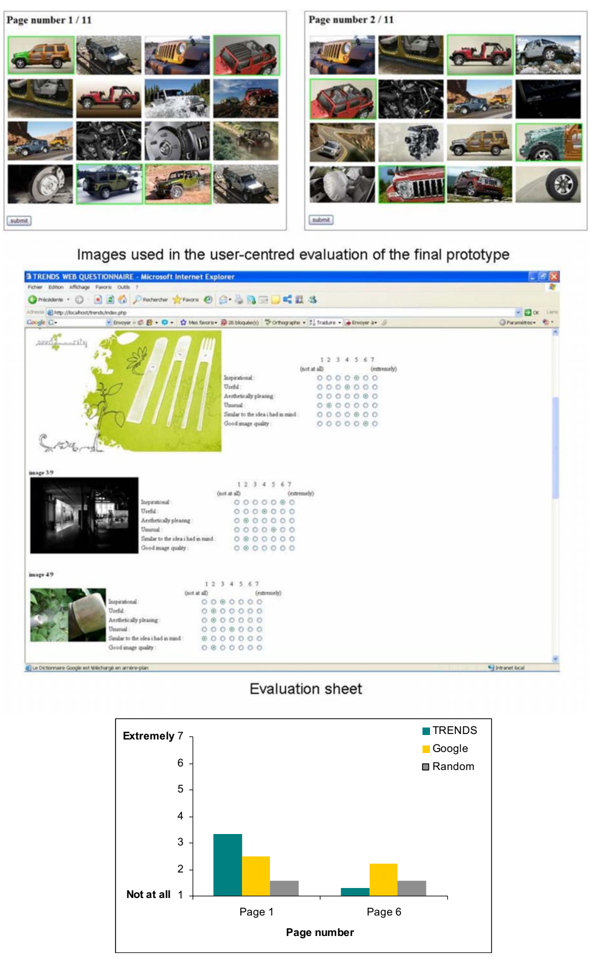


Fig. 8 User-centered evaluation: materials, evaluation sheet, and results (TRENDS, 2010)

Acknowledgment

The authors are grateful to their academic and industrial collaborators Stile Bertone, Italy, Fiat, Italy, Pertimm, France, INRIA, France, Robotiker, Spain, Leeds University, UK, and the European Community for funding the research described in this paper.

The authors also wish to acknowledge the contribution of Dr. Qiao Tang who created OntoRo, and Mr. E. Fadzli and Mr. I. Stankov who developed its web interface.

References

[1] Eckert, C. M., Stacey, M. K., and Clarkson, P. J., 2000, "Algorithms and Inspirations: Creative Reuse of Design Experience," Greenwich 2000 International Symposium: Digital Creativity, University of Greenwich, London, pp. 1–10.

[2] Westerman, S. J., Kaur, S., Mougenot, C., Sourbe, L., and Bouchard, C., 2007, "The Impact of Computer-Based Support on Product Designers' Search for Inspirational Materials," Proceedings of the Third I*PROMS International Conference, Cardiff, UK, Jul. 2–13, pp. 581–586.

[3] Ansburg, P. I., and Hill, K., 2003, "Creative and Analytic Thinkers Differ in Their Use of Attentional Resources," PAID, Vol. 34.

[4] Ford, N., 1999, "Information Retrieval and Creativity: Towards Support for the Original Thinker," Journal of Documentation, **55**(5), pp. 528–542.

[5] Eckert, C., and Stacey, M. K., 2000, "Sources of Inspiration: A Language of Design," Des. Stud., **21**, pp. 99–112.

[6] Bouchard, C., 1997, "Modelization of the Car Design Process," Ph.D. thesis, ENSAM, Paris, France.

[7] Mougenot, C., Bouchard, C., and Aoussat, A., 2006, "Fostering Innovation in Early Design Stage: A Study of Inspirational Process in Car-Design Companies," Wonderground—Design Research Society International Conference, Lisbon, Nov. 1–4.

[8] Lucero, A., Aliakseyeu, D., and Martens, J.-B., 2007, "Augmenting Mood Boards: Flexible and Intuitive Interaction in the Context of the Design Studio," Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 147–154.

[9] McDonagh, D., and Denton, H., 2005, "Exploring the Degree to Which Individual Students Share a Common Perception of Specific Trend Boards: Observations Relating to Teaching, Learning and Team-Based Design," Des. Stud., **26**, pp. 35–53.

[10] Gick, M. L., and Holyoak, K. J., 1980, "Analogical Problem Solving," Cogn. Psychol., **12**, pp. 306–355.

[11] Gomes, P., Seco, N., Pereira, F. C., Paiva, P., Carreiro, P., Ferreira, J. L., and Bento, C., 2006, "The Importance of Retrieval in Creating Design Analogies," Knowledge-Based Systems, **19**, pp. 480–488.

[12] Lai, I.-C., and Chang, T.-W., 2006, "A Distributed Linking System for Supporting Idea Association During the Conceptual Design Stage," Des. Stud., **27**, pp. 685–710.

[13] Bouchard, C., Aoussat, A., and Duchamp, R., 2006, "Role of Sketching in Conceptual Design of Car Styling," Journal of Design Research, **5**(1), pp. 116–148.

[14] Bonnardel, N., 2000, "Towards Understanding and Supporting Creativity in Design: Analogies in a Constrained Cognitive Environment," Knowledge-Based Systems, **13**, pp. 505–513.

[15] Mello, P., Storari, S., and Valli, B., 2008, "A Knowledge-Based System for Fashion Trend Forecasting," Lecture Notes in Artificial Intelligence, **5027**, pp. 425–434.

[16] Westerman, S. J., and Kaur, S., 2007, "Creative Industrial Design and Computer-Based Image Retrieval," The Role of Aesthetics and Affect, Lecture Notes in Computer Science, **4738**, pp. 618–629.

[17] Vijaykumar, G., Chakrabarti, A., 2008, "Understanding the Knowledge Needs of Designers During Design Process in Industry," ASME J. Comput. Inf. Sci. Eng., **8**, p. 011004.

[18] Cox, I. I., Miller, M. L., Minka, T. P., Papathomas, T. V., and Yianilos, P. N., 2000, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments," IEE Trans. Image Process., **9**(1), pp. 20–37.

[19] Kuroda, K., and Hagiwara, M., 2002, "An Image Retrieval System by Impression Words and Specific Objects Names—IRIS," Neurocomputing, **43**, pp. 259–276.

[20] Gusikhin, O., Rychtyckyj, N., and Filev, D., 2007, "Intelligent Systems in the Automotive Industry: Applications and Trends," Knowledge Inf. Syst., **12**(2), pp. 147–168.

[21] Tsai, C.-F., 2007, "A Review of Image Retrieval Methods for Digital Cultural Heritage Resources," Online Information Review, **31**(2), pp. 185–198.

[22] Datta, R., Joshi, D., Li, J., and Wang, J. Z., 2008, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Comput. Surv., **40**(2), pp. 1–60.

[23] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R., 2000, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Anal. Mach. Intell., **22**(12), pp. 1349–1380.

[24] Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y., 2007, "A Survey of Content-Based Image Retrieval With High-Level Semantics," Pattern Recogn., **40**, pp. 262–282.

[25] Ferecatu, M., Boujemaa, N., and Crucianu, M., 2008, "Semantic Interactive Image Retrieval Combining Visual and Conceptual Content Description," J. ACM, **13**(5–6), pp. 309–322.
[26] Enser, P. G. B., Sandom, C. J., Hare, J. S., and Lewis, P. H., 2007, "Facing the Reality of Semantic Image Retrieval," Journal of Documentation, **63**(4), pp. 465–481.
[27] Hyvonen, E., Styrman, A., and Saarela, S., 2002, "Ontology-Based Image Retrieval," *Towards the Semantic Web and Web Services. Proceedings of the XML Finland 2002 Conference*, E. Hyvonen and M. Klemettinen, eds., HIIT, Helsinki, Finland, pp. 15–27.
[28] Kiryakov, A., Popov, B., Terziev, I., Manov, D., and Ognyanoff, D., 2004, "Semantic Annotation, Indexing, and Retrieval," Web Semantics: Science, Services and Agents on the World Wide Web, **2**(1), pp. 49–79.
[29] Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, K., McCurley, S., Rajagopalan, S., Tomkins, A., Tomlin, J. A., and Zien, J. Y., 2003, "A Case for Automated Large-Scale Semantic Annotation," Web Semantics: Science, Services and Agents on the World Wide Web, **1**(1), pp. 115–132.
[30] Corcho, O., 2006, "Ontology Based Document Annotation, Trends and Open Research Problems, International Journal of Metadata," Semantics and Ontologies, **1**(1), pp. 47–57.
[31] Benjamins, V. R., Contreras, J., Blazquez, M., Dodero, J. M., Garcia, A., Navas, E., Hernandez, F., and Wer, C., 2004, "Cultural Heritage and the Semantic Web," Lecture Notes in Computer Science, **3053**, pp. 433–444.
[32] Köhler, J., Philippi, S., Specht, M., and Rüegg, A., 2006, "Ontology Based Text Indexing and Querying for the Semantic Web," Knowledge-Based Systems, **19**(8), pp. 744–754.
[33] Setchi, R., and Tang, Q., 2007, "Concept Indexing Using Ontology and Supervised Machine Learning, Transactions on Engineering," Transactions on Engineering, Computing and Technology, **19**, pp. 221–226.
[34] Setchi, R., and Tang, Q., 2007, "Semantic-Based Representation of Content Using Concept Indexing," Proceedings of the Third I*PROMS International Conference, Cardiff, UK, Jul. 2–13, pp. 611–618.
[35] Song, M., Song, I.-Y., Hu, X., and Allen, R. B., 2007, "Integration of Association Rules and Ontologies for Semantic Query Expansion," Data Knowl. Eng., **63**, pp. 63–75.
[36] Conesa, J., Storey, V. C., and Sugumaran, V., 2008, "Improving Web-Query Processing Through Semantic Knowledge," Data Knowl. Eng., **66**, pp. 18–34.
[37] Gene Ontology Consortium, 2001, "Creating the Gene Ontology Resource: Design and Implementation," Genome Res., **11**(8), pp. 1425–1433.
[38] WordNet, URL http://wordnet.princeton.edu/, last accessed 27/06/10.
[39] OpenCYC, URL, http://sw.opencyc.org/, last accessed 27/06/10.
[40] VIKEF, http://www.vikef.net/
[41] AIM@SHAPE: http://www.aimatshape.net/
[42] WIDE, http://www.ist-wide.info/
[43] X-MEDIA, http://www.x-media-project.org
[44] Bouchard, C., Mougenot, C., Omhover, J. F., Setchi, R., and Aoussat, A., 2007, "Building a Domain Ontology for Designers, Towards a Kansei Based Ontology," Proceedings of the Third I*PROMS International Conference, Cardiff, UK, July 2–13, pp. 587–592.
[45] Bouchard, C., and Aoussat, A., 2003, "Design Process Perceived as an Information Process to Enhance the Introduction of New Tools," Int. J. Veh. Des., **31**(2), pp. 162–175.
[46] Büsher, M., Frielaender, V., Hodgson, E., Rank, S., and Shapiro, D., 2004, "Design on Objects: Imaginative Practice, Aesthetic Categorisation, and the Design of Multimedia Archiving Support," Digital Creativity.
[47] Mougenot, C., Bouchard, C., and Aoussat, A., 2007. "An Experimental Study of Designers' Cognitive Activity in Design Information Phase," ICED 2007, 16th International Conference on Engineering Design, Paris, Aug. 28–3.
[48] Valette Florence, P., 1994, "Introduction a L'analyse des Chaînages Cognitifs," Recherche et Application en Marketing, **9**(1), pp. 93–118.
[49] Davidson, E., ed., 2003, *Roget's Thesaurus of English Words and Phrases*, Penguin, UK.
[50] Jarmasz, M., and Szpakowicz, S., 2001. "Roget's Thesaurus: A Lexical Resource to Treasure," Proceedings of the NAACL WordNet and Other Lexical Resources Workshop, Pittsburgh, pp. 186–188.
[51] Cassidy, P., 2000, "An Investigation of the Semantic Relations in the Roget's Thesaurus: Preliminary Results," CICLing-2000: Conference on Intelligent Text Processing and Computational Linguistics, A. Gelbukh ed., IPN Publishing House, Mexico City, Mexico, pp. 181–204; see http://www.cicling.org/2000/.
[52] Morris, J., and Hirst, G., 2004, "Non-Classical Lexical Semantic Relations," Workshop on Computational Lexical Semantics, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp. 46–51.
[53] Tang, Q., 2006, "Knowledge Management Using Machine Learning, Natural Language Processing and Ontology," Ph.D. thesis, Trevithick Library, Cardiff University, UK.
[54] OntoRo, URL http://kes.engin.cf.ac.uk/ontoro_x/, last accessed 27/06/10.
[55] Salton, G., and Buckley, C., 1990, "Improving Retrieval Performance by Relevance Feedback," J. Am. Soc. Inf. Sci., **41**(4), pp. 288–297.
[56] Gliozzo, A., Strapparava, C., and Dagan, I., 2004, "Unsupervised and Supervised Exploitation of Semantic Domains in Lexical Disambiguation," Comput. Speech Lang., **18**(3), pp. 275–299.
[57] OntoTag, 2008, Ontology Tagger, TRENDS Report, URL www.trendsproject.org, last accessed 27/06/10.
[58] Copernic Desktop Search, URL http://www.copernic.com/en/store/, last accessed on 27/06/10.
[59] TRENDS, 2010, URL (http://www.trendsproject.org/) last accessed 27/06/10 (2010). TRENDS D7.2 Prototype 3 Testing Report.