

Carrier Model for Semantically Annotated Information Goods¹

Wolfgang Maass¹, Wernher Behrendt² and Aldo Gangemi³

¹ Furtwangen University
wolfgang.maass@hs-furtwangen.de

² Salzburg Research Forschungsgesellschaft
wernher.behrendt@salzburgresearch.at

³Institute of Cognitive Sciences and Technology (CNR)
gangemi@loa-cnr.it

Abstract

Sharing and trading of information goods requires efficient and effective means for search and evaluation of the content carried by the information goods. The hurdle to overcome is the information paradox according to which, information goods obtain qualities of experience goods. This article describes an approach how ontology-based semantic annotations of information goods might reduce this paradox. At the centre of this approach is a generic container model called knowledge content object (KCO).

1 Introduction

The WWW can be perceived as a huge information market where supply and demand meet. If content has sufficiently high value for actors representing the demand side, it will generate market prices. This kind of content is generally termed "paid content" as a special form of information goods and is viewed as a digital product ([Boll/Klas 2000; Choi et al 1997]). [Shapiro/Varian 1999] define the term *information good* very broadly. "Essentially, anything that can be digitised - encoded as a stream of bits - is information. [...] Baseball scores, books, databases, magazines, movies, music, stock quotes, and Web pages are all information goods". Based on the definition of [Choi et al. 1997] anything one can send and receive over the Internet has the potential to be a digital product. "Information is a primary example of a digital product, for example knowledge-based goods that can be digitised and transferred over a digital network". The term paid content is in this article used as the non-free sales and distribution of information-based content products.

Research on the economics of information distinguishes between *search products* and *experience products* [Bloom 1989]. Search products are goods or services for which

¹ Based on results of the IST-FP6 project METOKIS (<http://metokis.salzburgresearch.at/>) that has been co-funded by the European Commission.

the most essential attributes can easily be evaluated prior to a purchase and provide a basis for an informed buying decision because consumers can verify claims before purchase [Franke et al. 2004]. Society is very much accustomed to buying search products such as cars, houses and computers. Experience products are goods or services for which the cost to evaluate the most essential attributes is so high that direct experience is often the evaluation method with the lowest costs in terms of time, money, cognitive effort, or other resources ([Franke et al. 2004]). Because of the difficulty involved in evaluating claims for experience products, consumers will be more sceptical of claims for experience products in comparison with search products. As Varian points out [Varian 2000], information has properties that are likely to cause difficulties for market transactions: (1) *experience good*: information needs to be experienced before you know what it is, (2) *returns to scale*: high fixed production costs and almost no reproduction costs, and (3) *public good*: information goods are typically non-rival and sometimes non-excludable [Varian 2000]. These three properties are the main reasons why the traditional content industry struggles to find business models that can be used with digital media. Hence, enrichment of information goods with semantic annotations evaluable by consumers provides the basis for shifting information goods into the class of search products, which we would expect to facilitate consumers' buying decisions.

Throughout this article we will explore the potential for paid content, of semantic annotations that provide on the one hand a facility for interoperable markets for paid content and on the other hand a means for product self-descriptions, which has strong influence on consumer buying decisions. First, we will briefly discuss application situations from which we will derive requirements for the intended semantically annotated content structure and the underlying technical content carrier architecture. In subsequent sections, we will introduce a semantically enriched container model (KCO). KCOs are exchanged over a transmission infrastructure that is based on a generic content carrier architecture (KCCA) which enables the interoperation of heterogeneous content repositories. Finally, economic implications of this approach are discussed before we will summarise and give an outlook on future work.

2 Knowledge intensive information goods

An analysis of several hundred existing paid content business models ([Stahl/Maass 2003]) resulted in a classification of five central elements to which digital content has to respond during different phases of its life cycle [Maass et al 2004, Behrendt et al 2005].

1. *Content descriptions*: provide the propositional content that is announced by a digital content on an abstract level.
2. *Community descriptions*: information about the organisational structure (roles, rights, obligations and prohibitions) by which a content product can be used and information that influences trust such as certificates and brand name information.
3. *Business descriptions*: describes the business and legal requirements during information and negotiation phases. After a purchase, contracts will be enforced according to mutually agreed rights and obligations.
4. *Presentation descriptions*: describes the presentation modes to which the information of a content product can be adapted to by rendering and other application-specific means.

5. *Trust and Security descriptions*: content must be associated with some measure of trust for the end user, and for the content provider, there must be some security features which guarantee that the content will not be illegally copied or otherwise misappropriated.

Based on this conceptualisation the concept of a Knowledge Content Object (KCO) intends to provide a structure on the computational level that can be implemented in digital infrastructures. From the ontological viewpoint given by the foundational ontology DOLCE+ [Masolo et al. 2003] and its extensions (a modular library called DDPO [Gangemi et al. 2004]), a KCO is to be distinguished from the abstract concept of an information object that carries meaning on cognitive and abstract level, independent of any technical realisation.

2.1 An Ontology of Information Objects

The main distinctions in the reused ontologies, which are imported here, include:

- the topmost class is called *particular* (any entity)
- *objects* (e.g. a *dog*) and *events* (e.g. a *barking*) belong to disjoint classes
- *physical* (e.g. a *brick*) and *social* (e.g. a *contract*) objects belong to disjoint classes
- attributes of particulars (e.g. a *colour*, or a *spatio-temporal location*) are represented as *regions* within *quality spaces*, with a possible associated metrics
- social objects include *descriptions* (the public, communicable counterpart of agents' conceptualizations, including also *plans*), which can define *concepts* (e.g. the *role of customer*), encode (or be expressed by) *information objects* (a *sentence*, or a *music chart*), provide unification criteria for *collections* (a *group of people*), etc.
- concepts can be either *roles* played by objects, *tasks* executed during actions (e.g. a *door opening task*), etc.
- concepts from descriptions provide constraints for other particulars: if a configuration of particulars satisfies those constraints, then a *situation* emerges that satisfies the concepts' description (typical applications of constraint unification include *regulations*, *plans*, *social relationships*, etc.).

The previous distinctions are supported by a large axiomatization.² Here, we concentrate on their application as a foundation for KCO implementation and deployment.

For example, a usage context of a content object may require us to talk about the digital reproduction of a painting that is owned by an institution, and such institution is willing to commercialize the reproduction at certain conditions that include differentiation for users, pricing, regulations to be followed, inclusion of content metadata, explanations, interpretations, ways of rendering it, etc. This context is complex, and requires a subtle differentiation of the various entity types involved in it.

According to DDPO, a content (information) transferred in any modality is a kind of social object called *Information Object* (IO). Information objects are *spatio-temporal reifi-*

² The full OWL axiomatization of DOLCE, DnS, DDPO, IO, etc. can be downloaded from: <http://dolce.semanticweb.org>.

cations of pure (abstract) information as described e.g. in Shannon's communication theory, hence they are assumed to be in time, and realized (materialized) by some entity. Information objects are the core notion of a *semiotic ontology design pattern*, which employs typical *semiotic relations* (c.f. [Gangemi et al. 2004]).

We present the axiomatization of KCOs in OWL abstract syntax. We firstly present the definition of `DnS:information-object`, which encodes the basic axioms of an ontology of semiotics extending the basic DOLCE ontology:

```
Class(DnS:information-object complete
  intersectionOf(
    DOLCE:social-object
    restriction(DnS:about allValuesFrom(DOLCE:particular))
    restriction(DnS:realized-by someValuesFrom(DOLCE:information-realization))
    restriction(DnS:interpreted-by allValuesFrom(Actions:agent))
    restriction(DnS:expresses allValuesFrom(DnS:description))
    restriction(DnS:ordered-by someValuesFrom(DnS:information-encoding-system))))
```

The definition says that information objects are necessarily *encoded* by some information encoding system, must be *realized* by some particular, can *express* a description, and, if that description is satisfied by a situation, can be *about* that situation, or some entity in its setting and can be *interpreted* by agents that can conceive the description expressed by said IOs.

For example, Barry Levinson's movie "Wag the Dog" is an IO, it is *ordered* by modern American English language (the information encoding system), is *realized* by, e.g., an DivX copy, *expresses* a certain plot on the media industry and politics and its related meaning, is *interpreted* by an agent in the role of a spectator with an average knowledge on making movies and election campaigns, and it is *about* certain entities and facts.

These *semiotic* relations constitute a typical *ontology design pattern*, so that any composition of relations can be built starting from any node in the pattern or in an application of the pattern.

3 Semantic Modelling of Knowledge Content Objects (KCO)

The notion of a knowledge content object is based on business requirements (see section 1), and builds upon previous approaches to multimedia and hypermedia document models. Related work includes [Boll/Klas 2001, van Ossenbruggen 2001, Zillner et al 2004]. The strength of KCOs lies in the combination of business- and domain-specific semantics (cf. section 1) that are tied into DDPO (cf. section 2).

Derived from the analysis of business models and paid content products, we have developed a KCO structure consisting of six facets. Several of these facets are subdivided into further KCO sub-facets. At the "atomic" level, it is intended that each of the leaf elements is associated with well-defined operational semantics, in order to enable organisations to quickly deploy KCOs as part of their information infrastructure.

<i>Facets</i>	<i>Elements</i>	<i>Short Description</i>
Content Description	Multimedia Characterization	How to access the content. This description includes information about the Content Format, Encoding, Storage Location, ... Depending on the domain, the Media Format Description of MPEG7 ³ or the content level description as defined in the NewsML ⁴ standard are good examples of content descriptions.
	Content Classification	How to classify the content: Keywords and concepts assigned to the content object based on a classification schema. Dublin Core or LOM are such classification schemas. The ITCP thematic thesaurus ⁵ and the ICON Class classification system ⁶ are well known controlled vocabularies in their domains.
	Propositional Description	What the content means. This semantic description is about the subject of the Content Object and not about the Content object itself.
Presentation Description	Spatio-temporal rendition	How to present the content: Description of how the content (and the Knowledge) of the KCO is presented to users. Presentation includes the rendering, rendition as well as interaction models. The SMIL ⁷ standard is clearly related to this facet.
	Interaction-based rendition	
Community Description	User task	How the Content is typically used: Description of Plans, Tasks, Roles and Goals in the context of a community which uses KCOs
	User community	
	Usage history	How the Content was used/changed: List of actions performed with the KCO during its lifecycle
Business Description	Negotiation protocol	How to trade the content: Business processes define special plans (negotiation protocols) and roles (auctioneer, seller, buyer,...) related to some business activity. This facet can be viewed as a specialisation of the community facet.
	Pricing scheme	
	Contract information	
Trust & Security	none	How to protect the content:
Self-description	none	How to understand the KCO: Specification of the (inner) structure of the KCO (e.g., active facets, ontologies used) in machine-interpretable form.

³ <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>

⁴ <http://www.newsml.org/pages/index.php>

⁵ <http://www.iptc.org/NewsCodes/>

⁶ <http://www.iconclass.nl/>

⁷ <http://www.w3.org/TR/2005/REC-SMIL2-20050107/>

The first facet is what we call the *content description*. The KCO carries a list of media references which are intended to point to real media files. So the collection of these referenced media files is actually, the full intended content of the KCO. In order to make this content accessible for machines, we provide a simple referential mechanism to associate arbitrary logic descriptions to the media files. For this, the propositional content facet is linked to a domain ontology which represents the universe of discourse for all content descriptions of this KCO. We also include a facet element which allows the use of existing classification and meta data schemes that need not necessarily be linked to the DOLCE foundational ontology. Its purpose is to enable the migration of media and their associated metadata into the KCO structure. This provides a basis for *ontological alignment* at a later stage.

The second facet is the specification of time-based spatial *presentation of, and interaction* with, complex content. Given some media tokens, we specify on one or more temporal “tracks” which describe *when* the associated media data will be rendered, and *where* they will be rendered (in terms of spatial arrangements). The second facet element deals with *interaction* and *dialogue*. Here, the semantic annotation specifies whether the presentation is entirely pre-programmed, whether it is entirely open (e.g. web based navigation) or whether it follows some dialogue pattern where humans and system take conversational turns in order to navigate the knowledge/information structure. This description defines one or more discourse structures that can be associated with the content for its rendering.

The third facet, *community description*, describes the context in which a content can be used. This covers three facet elements: *user tasks*, which are formally described by reference to an ontology of plans and tasks; *user community*, which describes the situations with corresponding roles (rights and obligations) that users would take in order to manipulate or consume the content; and *usage history*, keeping traces of previous use in order to support workflow systems as well as collaborative filtering systems. The latter can be achieved by keeping track of user data when the KCO is being “touched” by that user. Depending on legal and contractual aspects of this facet, the filtering may be more or less anonymous.

The fourth facet, *business description*, contains a specification of the business semantics associated with the KCO. This comprises the facet element negotiation protocol which describes the business scripts by which a contract is being negotiated. A negotiation protocol is described as a DDPO *plan* that can be represented and processed in OWL-DL format [Gangemi et al 2004]. The pricing scheme is used for restricting the price policies that can be applied during the negotiation. It is grounded in DDPO as the *regulation* concept. In the simple case of a fixed-price scheme, the negotiation is reduced to a simple over-the-counter (OTC) purchase. The pricing scheme is required for price differentiation strategies that are defined by the seller on the basis of a differentiating factor such as age, quantity discount or date of content origin (see [Stahl/Maass 2004, van Ossenbruggen et al. 2001]). The resulting contract is also a plan in the DDPO ontology that describes the situation in which agentive roles can be taken by agents and act by using described tasks. We distinguish between pre-existing content and prospective content. Pre-existing content is already available at contracting time while prospective content is produced during execution time of the contract.

The fifth facet, *trust and security*, is currently deemed out of the scope of our project although we acknowledge the need for inclusion of the issue, in the overall framework of METOKIS. The framework is rounded off with the sixth facet, *self-description*.

The KCO model is translated into an OWL ontology, which specializes the IO branching of DDPO foundational ontology (ref. [Behrendt et al. 2005]).

4 Electronic Markets based on KCCA enabled infrastructures

In order to make use of the semantic richness that can be expressed with KCOs we need an infrastructure whose components support the functionality afforded by the KCO. The Knowledge Content Carrier Architecture (KCCA) does this in the shape of services which are logically clustered by KCCA's components arranged in a 3-tier architecture. This gives rise to the following structural core components: (1) KCO Service API - offering the functions described by the facets in table 1, (2) KCCA Registry and Manager - managing a federation of KCO-aware nodes, (3) KCTP Service - a protocol to exchange service requests across KCCA nodes and (4) KCCA Profiles - Services for the wrapping and integration of data sources.

One of the assumptions of our work is that eventually, most information systems will make use of two further components: firstly, reasoning services based on ontologies and secondly, a task execution environment that will support the definition and execution of flexible workflows. KCOs are designed to support such an architecture through their content description (this is where reasoning services can access the KCO) and through their community description (describing the tasks for which this KCO is useful and the roles of actors that would do the tasks).

We envisage future publishing environments to use an integrated framework consisting of the components described. This will leave the application builder to focus on application and domain specific adaptations, and on the tailoring of the presentation /interaction layer to the needs of the customer.

The following architectural overview shows the full picture combining KCCA components, reasoning and task execution environment, as well as domain specific adaptations and the application layer.

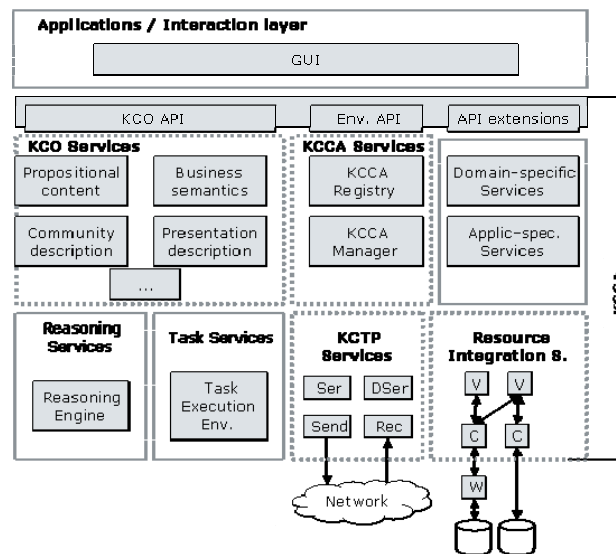


Fig. 1: Knowledge Content Carrier Architecture (KCCA).

The KCO services offer access to the operational semantics of the KCO facets. The KCCA Registry and Manager component keeps track of how a federation of KCO aware information systems is set up. The KCCA environment keeps information about information sources, wrappers and maintains state in user sessions that may span requests and transactions across the federation. The KCTP Services define a stateful protocol⁸ that allows communication between KCCA nodes by exchanging serialised RDF graphs. The KCCA Integration Services give assistance in binding non-KCCA resources to a KCO aware system. This is done by a two-stage mapping process. The external information source is first mapped into an equivalent RDF schema which we call "context profile". This can be a "naive" mapping to RDF. Next, a view is defined over the context profile and this view is made KCO compliant ("view profile"). The provider of an external information source needs to write a wrapper which provides the context profile for the resource. The KCCA integrator uses the context profile to create the view profile.

5 Economic implications

From an economic point of view five hypotheses can be derived from the KCO carrier model for information goods: (1) reduction of transaction costs, (2) increase of perceived value of high-quality information goods, (3) increased competition in markets for information goods and induced price reduction, (4) reduction of operational costs by reuse of knowledge, and (5) positive influence on the consumer's decision making. These hypotheses are discussed in turn.

KCO-contained information goods are designed to be shared and traded by electronic markets because they can leverage the standardised format of KCOs. Any kind of information good is wrapped into a homogeneous format that can be carried by any KCTP-supporting infrastructure. Facet information can be perceived as plug-ins delivered by providers. The facet information provided can be used by consumers of electronic markets to reduce their information costs [Bakos 1991, Malone 1987].

As a consequence of the first hypothesis, users will be able to evaluate the value of information goods for less cost which should increase the diffusion and adoption particularly of complex information goods [Maass/Stahl 2006]. In general it can be assumed that high-quality information goods will extend their reach and life time, i.e. increase of revenue for paid contents.

If semantically annotated information goods reduce transaction costs and increase their reach, the size of markets should increase, too. Because KCO-contained information goods provide a modular but sufficiently complete structure, consumers gain improved means for product comparison. Obviously, providers will reduce this effect by differentiation of their provided semantic information. Commoditised information goods, such as stock or weather information, are likely to become target of intensified competition [Bakos 1991, Malone 1987].

If hypotheses 1 to 3 will hold, then it can be assumed that consumers will be more willing to use electronic markets as knowledge channels, i.e., they outsource knowledge competencies to external providers. Organising and archiving knowledge coded in information sources is handed over to external services if it is more effective and efficient

⁸ The protocol is FIPA-based and can be implemented on top of SOAP or http.

than internal provisioning [Apostolou et al. 2005]. Container models, such as KCO, enable this outsourcing process for complex information goods. This process starts with simple information goods such as bookmarks (e.g. del.icio.us or shadows.com), pictures (flickr.com, bubbleshare.com), videos (videoegg.com and jumpcut.com), feeds (feedblendr.com), wikis, blogs and email conversations (9cays.com) but started even with complex information goods such as software repositories (sourceforge.net, colab.net).

Finally, recent results in researching consumer behaviour indicate that information given by electronic recommender systems positively influence consumers' decision to buy a particular good and furthermore reduce the amount of alternative goods considered [Häubl/Murray 2006]. In particular facets 1 to 5 contain information which is likely to be used by consumers for making their decision. However, empirical tests will be required to evaluate this hypothesis.

6 Summary and open issues

We have discussed how information goods can be semantically described by and included into a KCO container model. Other approaches such as Adobe's XMP⁹ follow similar goals. It is most likely that several other architectures will appear before a solution is accepted by a large audience. The KCO/KCCA approach paves the ground and explores application areas and technological solutions.

In the future we will investigate how KCO-embedded information goods can be used to generate new kinds of information goods. This can be either done manually as part of an editorial process or automatically based on plans and situations. For instance, new information goods can be aggregated based on user preferences, willingness-to-pay or user locations to name a few. How and by which means KCO facet information can be processed is part of future research.

References

- Apostolou, D., Mentzas, G., Abecker, A., and Maass, W. (2005). Challenges and directions in knowledge asset trading. *Intelligent Systems in Accounting, Finance and Management*, 1-16.
- Bakos, J. (1991): A Strategic Analysis of Electronic Marketplaces, *MISQ*, 15(3), S. 295-310, 1991.
- Behrendt, W., Gangemi, A., Maass, W., & Westenthaler, R. (2005). Towards an Ontology-based Distributed Architecture for Paid Content.. In Gómez-Pérez, A., & Euzenat, J. . (Eds.), Springer (pp. pp. 257). Berlin.
- Bloom, P. (1989): A Decision Model for Prioritizing and Addressing Consumer Information Problems. *Journal of Public Policy & Marketing*, 8 (1). 1989, 161-180.
- Boll, S. and Klas, W. (2001): ZYX - A Multimedia Document Model for Reuse and Adaptation. In: *IEEE Transactions on Knowledge and Data Engineering*, 13(3), 2001, pp. 361-382.

⁹ <http://www.adobe.com/products/xmp/>

- Choi, S.Y., Stahl, D.O. and Whinston, A.B. (1997): *The Economics of Electronic Commerce*. Macmillan Technical Publishing, Indianapolis, 1997.
- Franke, G., Huhmann, B. and Mothersbaugh, D. (2004): Information Content and Consumer Readership of Print Ads: A Comparison of Search and Experience Products. *Journal of the Academy of Marketing Science*, 32 (1). 2004, 20-31.
- Gangemi, A., Borgo, S., Catenacci, C. and Lehmann, J. (2004): Task Taxonomies for Knowledge Content, METOKIS Deliverable, D07, 2004.
- Häubl, Gerald and Kyle Murray (2006), "The 'Double Agent' – Benefits and Pitfalls of an Electronic Agent's Personalized Product Recommendations," *Sloan Management Review*, 47, 3, forthcoming.
- Maass, W., & Stahl, F. (2006). Adoption and Diffusion in Electronic Markets: An Empirical Analysis of the German Paid Content Market. *Electronic Market*, forthcoming.
- Maass, W., Goyal, S., and Behrendt, W. (2004). Knowledge Content Objects and a Knowledge Content Carrier Infrastructure for ambient knowledge and media aware content systems. In *Knowledge-based Media Analysis* (pp. 449-456). London: QMUL.
- Malone, T. (1987): Modelling Coordination in Organizations and Markets, *Management Science*, 33, 1987, 1317-1332.
- Masolo, C., Borgo, S., A., G., Guarino, N. and Oltramari, A. (2003): *The WonderWeb Library of Foundational Ontologies*, 2003.
- Reich, S., Behrendt, W., Eichinger, C. and M., Schaller, M. (2000): Document Models for Navigating Digital Libraries. in *International Conference on Digital Libraries*, (Kyoto, 2000), 277-284.
- Shapiro, C. and Varian, H.R. (1999): *Information rules - A Strategic Guide to the Network Economy*. Harvard Business School Press, 1999.
- Stahl, F., F., S. and Maass, W. (2004): Paid Content - Paid Services: Analysis of the German Market and Success Factors of 280 Business Models, =mcminstitute, University of St. Gallen, St. Gallen, 2004, 163.
- van Ossenbruggen, J., Geurts, J., Cornelissen, F., Rutledge, L. and Hardman, L., (2001): Towards Second and Third Generation Web-Based Multimedia. in *The Tenth International World Wide Web Conference*, (Hong Kong, 2001), 479-488.
- Varian, H.R. (2000): Markets for Information Goods. in *Monetary Policy in a World of Knowledge-Based Growth, Quality Change, and Uncertain Measurement*, 2000.
- Zillner, S., Westermann, U. and Winiwarter, W. (2004): EMMA - A Query Algebra for Enhanced Multimedia Meta Objects. *CoopIS/DOA/ODBASE*, 2. 1030-1049.