006550
FOFIS
Formal Ontology for Information Systems

EIF

Deliverable 2
Ontology of Information Objects

Period Number:1 Due date of deliverable: 17.03.2006

Period covered: from 1.02.2005 to 31.01.2006 Date of preparation: 5.03.2006

Date of submission: 10.03.2006

Start date of project: 1.02.2005 Duration: 12 months

Project coordinator name: Nicola Guarino
Project coordinator organisation name: Istituto di Scienze e Tecnologie della Cognizione del Consiglio Nazionale delle Ricerche

Organisation name of lead contractor
for this deliverable: Istituto di Scienze e Tecnologie della Cognizione del Consiglio Nazionale delle Ricerche

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)
Dissemination in level PU

Contents

1 Introduction

In this deliverable I provide a formal theory of information objects construed as artefacts of special sort, or, to be more modest, I provide a formal ontology of documents. Among various general theories of the latter (see e.g. [7], [9], [10], [13], [15], [16], [18], [26]) I have chosen the theory of document genres, whose basic assumptions justify such construal.

The idea of applying the notion of document genre in information systems is now widely recognised. There is a number of theoretical and practical studies in which documents are represented in terms of their genres. The Digital Document Track of the annual Hawaii International Conference on System Science has become an established forum for presenting these results. The specific domains of application include information and document retrieval, metadata schemas, computer-mediated communication, electronic data management, and computer-supported collaborative work.

Nonetheless, the very notion of genre is unstable and the conceptual divergences between different theories thereof are substantial. For example, it is debatable whether we should represent a genre by means of pairs <substance, form>, as suggested in [28], or triples <substance, form, functionality> ([14]) or quadruples ([22]). Some even deny that all different kinds of genres may be represented in a uniform way ([5]). There is no agreement on what kinds of genres there are and how one may organise them in a taxonomy. In particular, the theoretical status of the so called cybergenres is disputed.

I believe that at least some of these issues may become much more transparent if we specify the ontological commitments of the genre discourse. It is usually believed in Knowledge Representation that a clear picture of the ontology that stands behind a given vocabulary/database schema/taxonomy/discourse model/... may contribute both to the theoretical adequacy of the latter and to its practical applicability or efficiency. The aim of this deliverable is to construct a precise ontological framework in which the notion of genre may be defined in such a way that we could understand what "ontological price" we need to pay for document genres. The framework in question should clarify what entities we need to acknowledge in order for our talk about document genres not to be void. Since this is the first ontological inquiry into the domain of the genre discourse, the results are to be interpreted as partial and preliminary.

2 Genres in organisational communication

The notion of genre I focus on here originates in the theory of organisational communication. J. Yates and W. Orlikowski define it in the following way:

> A genre of organizational communication (e.g. a recommendation letter or a proposal) is a typified communicative action invoked in response to a recurrent situation. The recurrent situation or socially defined need includes the history and nature of established practises, social relations, and communication media within organizations (e.g. a request for a recommandation letter assumes the existence of employment procedures that include the evaluation and documentation of prior performance [...]). ([28], p. 301)[1]

A genre is claimed to consist of substance and form. The former aspect encompasses the topics and needs addressed in a given act of communication and the purposes of undertaking of such act. The latter is claimed to be related to the physical features of the document. [28] mentions in this

[1] For different definitions of document genre see e.g. [20], p. 25; [24], p. 58; [2], p. 4.

context its structural features, the medium in which the document is storaged, and the respective language system. Genres are dynamic entities: they are enacted, reproduced, and transformed. [28] shows how to describe such processes by means of the notion of social rule taken from the structuration theory of social institutions (cf. [8]). A genre rule associates the form and substance of a given genre with certain recurrent situations.

> For example, in the case of the bussiness letter, which is invoked in recurrent situations requiring documented communication outside the organization, the genre rules for substance specify that the letter pertain to a bussiness interaction with an external party, and the genre rules for form specify an inside address, salutation, complimentary close, and correct, relatively formal language. ([28], p. 302)

The relation between a genre and its genre rules is not very tight:

> A particular instance of a genre need not draw on all the rules constituting that genre. For example, a meeting need not include minutes or a formal agenda for it to be recognizable as a meeting. Enough distinctive genre rules, however, must be invoked for the communicative action to be identified - within the relevant social community - as an instance of a certain genre. A chance encounter of three people at the water cooler, which is not preplanned and lacks formal structuring devices, would not usually be considered as a meeting. ([28], p. 302-303)

By applying genre rules in recurrent social situations, individual agents maintain genres; by modifying genre rules, they modify old genres or create new ones. The main factors in genre modification comprise new social, technological, and economic situations, and new media developments.

[21] introduces the notion of genre repertoire. A *genre repertoire* for a community is a set of genres routinely enacted within this community. Any genre repertoire is described by two parameters: its composition and its use. The former coincides with the content of the genre repertoire; the latter assigns to each genre in the repertoire the frequency with which this genre is enacted. A coordinated sequence of genres enacted by members of a particular organisation constitutes a *genre system*. For instance, the genre system of balloting was identified as consisting of three genres: the ballot form issued by the group coordinator, the ballot replies generated by the group members, and the ballot results. ([30], p. 51)

Any (sufficiently capacious) collection of genres may be meaningfully ordered with respect to their generality. Yates and Orlikowski emphasise that any subsumption hierarchy of genres is relative to a social context.

> [...] the positive recommendation letter could be viewed as a subgenre of the recommendation letter, which a subgenre of the business letter. [...] In the contemporary American climate [...] a situation may be emerging in which almost all recommendation letters are positive and, thus, the three nested genres can be collapsed into two genres. ([28], p. 303-304)

In a series of papers: [21], [30], [29], [11], Yates and Orlikowski showed that this theoretical framework is well-suited for empirical study of electronic-supported communication in real-world organisations. The genre discourse turned out to be a fruitful methodology also in web information retrieval as attested by [5],[6], [14], and [22].

3 Ontological commitments of the genre discourse

Speaking about ontological presuppositions of the genre discourse, we should distinguish between particular tokens of a certain genre and the type of this genre. The distinction between tokens and types may be characterised in terms of the relation of instantiation. Any particular token of a certain genre is said to instatiate the type of this document genre. For example, a particular job application instantiates the type of the job application genre. In what follows I will call any token (i.e. instance) of a document genre a *document*. Simlarly, any type of a document genre will be called a *genre*. I assume that both documents and genres are construed along the lines of the theory of Yates and Orlikowski as sketched above. However, I do not presuppose that this theory is the only plausible theory of genres or of documents.

I propose to articulate the genre discourse by means of the following primitive notions:

1. two basic general ontological categories of endurants and perdurants,
2. a specific relation of being a member of a community,
3. a complex general ontological category of situation-types,
4. a non-empty set $Time$ of time parameters (temporal moments or regions),
5. a specific ontological category of agents and three specific relations between agents' mental attitudes and situation-types,
6. two specific relations of being a part of, one of which is atemporal and the other is temporal,

In other words, I submit that the above categories (together with their short descriptions below) are sufficient ontological commitments of the genre theory of Yates and Orlikowski. I do not claim that they are necessary; still, I conjecture that it is improbable that one can provide a less ontologically demanding framework. Although these categories are assumed here to be primitive, in order to avoid (or decrease) confusion, I will briefly characterise some of them.

Endurants and perdurants. The notions of endurant and perdurant are understood here as usual. An *endurant* is an entity that is wholly present, i.e. whose all parts are present, at any time at which it exists. A *perdurant* is an entity that enfolds in time, i.e. for any time at which it exists, some of its parts are not present. How to draw a line between endurants and perdurants is a controversial isssue, however people, cars, and books are usually considered as endurants and people's lives, car races, and acts of reading are considered as perdurants. A set End will contain all endurants we need for a given genre discourse and a set $Perd$ will contain all relevant perdurants. What is not controversial is the claim that no endurant is a perdurant.

$$End \cap Perd = \emptyset. \tag{1}$$

In our formal ontology we need both endurants and perdurants because we saw above that some documents are endurants, e.g. a memo, but other are predurants, e.g. a meeting.

Communities and their members. According to the genre theory of Orlikowski and Yates, any document (and thereby any genre) is enacted, maintained, and transformed by and within a certain community. I will represent this aspect of the theory by introducing the relation of membership. The expression "x *in* y" is to mean that an endurant x is a member of a community y.

$$x \in Com \equiv \exists y \in End\ y\ in\ x. \tag{2}$$

Thereby I assume that communities are entities that do not change their membership trough time. If a genre x is enacted, maintained or transformed in a community y, I will say that x *comes from* y.

Situation-types, agents, and mental attitudes. The term "situation-type" is understood here as referring to such ontologically complex entities as that John is unemployed, that John's car first stopped and then burst into flames, and that Peter will steal John's book. More generally speaking, any entity to which somebody refers by means of a sentence will be called here a *situation-type*. The ontological category I have in mind here coincides with the category of situation-types as defined and used in [1] and the category of states of affairs as defined and used in [12]. The set of all situation-types that we need in the genre discourse will be denoted by the symbol "Sit". It is important to emphasise that in principle any situation-type may obtain at one moment (temporal region) and not obtain at another. For instance, that Andrea Merkel is a chancellor obtains in January 2006 and did not obtain in March 2004.

The notion of situation-type is used here to model the conditions under which and the purposes for which a document is created. Some of such conditions refer to the objective facts. For instance, given that an annual report is created periodically, the fact that we are now in such a period is an objective situation-type. Other conditions and all purposes are related to the subjective facts such as those that somebody entertains certain belief or desire, e.g. a ballot form is issued when someone desires information about the beliefs of certain people. I isolate within the set Sit a subset Sit_0 that contains the situation-types of the former kind. Let a set $Agt \subseteq End$ contain intentional agents, i.e. those endurants that are capable of bringing it about that situation-types obtain and capable of entertaining beliefs, desires, and intentions. Obviously,

$$x\ in\ y \rightarrow x \in Agt. \tag{3}$$

In order to include the subjective situations in Sit, I will use the following inductive definition:

$$Sit_{n+1} := Sit_n \cup \{< x, y >: x \in Agt \wedge y \in Sit_n\}. \tag{4}$$

$$Sit_\omega := \bigcup Sit_n. \tag{5}$$

The specific content of Sit may be established by one of the axioms of the form 6.

$$Sit := Sit_n. \tag{6}$$

Although different kinds of communities seemingly require different values of the parameter n, there seems to be two distinguished points: $n = 2$ and $n = \omega$. These points determine two different ways of modelling the notion of mutual belief, which is of crucial importance in any kind of theoretical reflection on social reality. The former point is related to the claim that we find e.g. in [25] on p. 41-51 to the effect that in most cases it is sufficient (and necessary) to define this

notion in terms of second-order beliefs. Briefly speaking, all members of a community mutually believe that p iff they all believe that p and they all believe that they all believe that p. The latter point is related to the iterative notion of mutual belief (e.g. [17], p. 52-60), which requires to this end n-order beliefs, for any $n \in \omega$. Briefly speaking, all members of a community mutually believe that p iff they all believe that p, they all believe that they all believe that p, and they all believe that they all believe that they all believe that p, ... Because it is highly improbable that any member of any real-world organisation that produces and uses documents entertains such "infinite" beliefs, I adopt the former notion, which in the present framework may be defined by 8. To this end, I first fix the value of the parameter n in 6 to be equal to 2. Next, I assume that all mental attitudes to which one is committed in his genre discourse may be defined in terms of beliefs ($Bel \subseteq Agt \times Sit$), desires ($Des \subseteq Agt \times Sit$), and intentions ($Int \subseteq Agt \times Sit$). "$< x, y > \in Bel$" stands for the expression "x believes that a situation-type y obtains". Analogously, I read the abbreviations "$< x, y > \in Des$" and "$< x, y > \in Int$". Consequently, I treat beliefs, desires, and intentions as situation-types. Among different possible assumptions concerning the relationships between beliefs, desires, and intentions (see e.g. [27], p. 99-102), I adopt the modest claim to the effect that intentions entail desires.

$$Int \subseteq Des. \tag{7}$$

$$\begin{aligned} &\text{A community } x \text{ has a mutual belief that } y \text{ obtains } \equiv \\ &\equiv \forall z(z \; in \; x \rightarrow < z, y > \in Bel \wedge < z, < z, y >> \in Bel). \end{aligned} \tag{8}$$

.

In what follows, I will need two auxiliary concepts defined by 9 and 10.

$$Ment\_Sit := Bel \cup Des \cup Int. \tag{9}$$

$$x \in Com \rightarrow Ment\_Sit(x) := \{< y, z > \in Ment\_Sit : y \; in \; x\}. \tag{10}$$

It should be obvious that no situation is neither an endurant nor a perdurant.

$$Sit \cap (End \cup Perd) = \emptyset. \tag{11}$$

I do not wish to take any stance on the issue whether communities are endurants or perdurants (or whether some are endurants and other are perdurants). Nevertheless, leaving this issue open, I claim that no community is a situation-type.

$$Com \cap Sit = \emptyset. \tag{12}$$

Parthood relations. Our two basic categories of endurants and perdurants need two relations of parthood. Since endurants may loose and gain (spatial) parts over time, speaking about their mereological structure, we need specify a temporal point of reference. On the other hand, since perdurants cannot loose or gain parts, we should describe their mereological structure from an atemporal point of view. This solution follows the distinction adopted in [19].

When we describe the mereological structure of a genre, we do not use the term of "part" in the sense of the standard mereological system of S. Lesniewski (see e.g. [4]). The reason for this claim is simple: such mereological theorems as the axiom of generalised sum, when applied to genres, postulate the existence of entities which are never mentioned in the genre descriptions. For

instance, all lists of parts that you find in the genre discourse are limited. You do not find therein such exotic entities as the mereological sum of the second chapter of a given book and the last word in the last chapter, although you can find chapters and words. Therefore, instead of modelling such mereological structures in terms of the standard mereology, I need other, less-demanding, notion of parthood. Among different weaker theories of parthood, I opt for a theory defined in [23]. To be more precise, I will borrow this theory for my temporal relation of parthood; as for its atemporal counterpart, I will simply strip this theory from its temporal indices. Let "$x \leqslant_t y$" mean that an endurant x is a part of an endurant y at t ($t, t_1, \ldots \in Time$). Let "$exist(x,t)$" mean that an endurant x exists at t. Definitions 13 and 14 introduce two auxiliary notions.

$$x <_t y \equiv x \leqslant_t y \wedge \neg y \leqslant_t x. \tag{13}$$

$$x \circ_t y \equiv \exists z(z \leqslant_t x \wedge z \leqslant_t y). \tag{14}$$

$$exist(x,t) \rightarrow x \leqslant_t x. \tag{15}$$

$$x \leqslant_t y \rightarrow exist(x,t) \wedge exist(y,t). \tag{16}$$

$$x \leqslant_t y \wedge y \leqslant_t z \rightarrow x \leqslant_t z. \tag{17}$$

$$x <_t y \rightarrow \exists z(z <_t y \wedge \neg z \circ_t x). \tag{18}$$

The atemporal notion of parthood for perdurants is defined by means of definitions 19 and 20, and axioms 21, 22, and 23.

$$x < y \equiv x \leqslant y \wedge \neg y \leqslant x. \tag{19}$$

$$x \circ y \equiv \exists z(z \leqslant x \wedge z \leqslant y). \tag{20}$$

$$x \leqslant x. \tag{21}$$

$$x \leqslant y \wedge y \leqslant z \rightarrow x \leqslant z. \tag{22}$$

$$x < y \rightarrow \exists z(z < y \wedge \neg z \circ x). \tag{23}$$

Besides, I add two constraints on the ontological categories of arguments of $\leqslant_t$ and $\leqslant$.

$$x \leqslant_t y \rightarrow x, y \in End. \tag{24}$$

$$x \leqslant y \rightarrow x, y \in Perd. \tag{25}$$

At the present stage of this theory, the precise strength of the mereological principles is not crucial. For instance, instead of the weak supplementation principle (i.e. 18 and 23) we can choose the strong supplementation principle (as suggested in [3], p. 39).

4 Towards a formal definition of genre

I will define a genre as a (set-theoretical) pair whose elements correspond to the informal definition from section 2 supplemented with the following extensions and modifications:

1. I will carefully distinguish between a document genre and a communication genre. The former is instantiated by documents that are endurants; the latter is instantiated by documents that are perdurants. Any document of a document genre will be called a *document in the strict sense*; any document of a communication genre will be called an *act of communication* or just a *communication*.

2. Since the description of the concept of genre that we find in [28] contains heterogeneous components, I will reorganise it by splitting the aspects of substance, form, and genre rule, and joining them into two elements: use and content.

3. The *use* element of a genre is to contain the recurrent situations in which the genre is referred to and the purposes for which it is referred to. The former aspect will be represented here by a set $Trigger$ of situation-types. $Trigger$ is to comprise all conditions that are necessary for production of a document of a given genre. Any element of $Trigger$ will be called a *trigger* both for the genre and for the documents of this genre. Because all triggers are situation-types, any document of a genre is associated with the same set of triggers. Similarly, the purpose aspect will be represented by a set $Purpose$ of situation-types. Each element of $Purpose$ will be called a *purpose* both of a given genre and of all documents of this genre.

4. Since any document is produced because of some mental attitude of some agent, at least one trigger for a document is a situation-type related to some mental attitude. Since any document is produced in order to evoke some mental attitude of some agent (to inform, to encourage, to request, etc.), at least one purpose of a document is a situation-type related to some mental attitude. Since any document is produced within some community, I assume that at least one trigger for or purpose of a genre from a community x, belongs to $Ment\_Sit(x)$.

5. The topics addressed by a genre and its form aspect will be united together by the notion of content. The *content* of a genre consists of the medium and the language of the genre. The former is to represent the medium and structure components of the form aspect from the theory of Yates and Orlikowski. The latter is to represent the language component of theirs and, to some extent, the topics addressed by the genre. The medium component of my concept of genre contains a set of genre supports and a relation among characteristic parts of these supports. A *support* for a genre is any document of this genre. This implies that any endurant or perdurant that was, is, or will be created in a given community as an instance of some genre, is treated as a support of this genre. Since supports are particular entities, each of them has its own mereological structure, which in the case of endurants may change over time. On the other hand, all documents of a given genre should share the same mereological pattern due to which they belong to the same genre. Consequently, I claim that for each document from a given genre, there exists a set of its parts, which will be called *characteristic* for this genre, such that a set of characteristic parts of any other document from this genre is isomorphic to the former set. Any characteristic part of a document contributes to the structural specificity of this document in so far as this specificity is determined by the genre to which this document belongs. Examples of such characteristic parts include paragraphs, titles, salutation lines, etc. In the case of documents in the strict sense, it seems obvious that only their essential parts may be characteristic. A part x of an endurant y is *essential* for y iff whenever x exists, it is a part of y.

$$x, y \in End \rightarrow [x \leqslant_{es} y \equiv \forall t\ (exist(y,t) \rightarrow x \leqslant_t y) \wedge \exists t\ exist(y,t)]. \tag{26}$$

   Therefore, my definition of characteristic part assumes that if x is a characteristic part of a document y of some genre, then x is essential for y.

6. The content of a genre will be represented as a pair $< Med, Lang >$, where a set Med characterises the medium aspect of the genre and a set $Lang$ characterises its linguistic dimension.

7. The medium of a document genre will be represented as a pair $< Supp, \leqslant_{ch} >$, where

   (a) $Supp \subseteq End$ is a non-empty set of supports of a given genre,

   (b) $\leqslant_{ch}$ is a subset of $\leqslant_{es}$ such that $\leqslant_{ch}$ is a partial order and

   $$\forall x, y \in Supp < P_{\leqslant_{ch}}(x), \leqslant_{ch} > \text{ is isomorphic to } < P_{\leqslant_{ch}}(y), \leqslant_{ch} >, \tag{27}$$

   where $P_{\leqslant_{ch}}(x) := \{y \in End : y \leqslant_{ch} x\}$.

8. The medium of a communication genre will be represented as a pair $< Supp, \leqslant_{ch} >$, where

   (a) $Supp \subseteq Perd$ is a non-empty set of supports of a given genre,

   (b) $\leqslant_{ch}$ is a non-empty subset of $\leqslant$ such that $\leqslant_{ch}$ is a partial order and condition 27 is satisfied, now $P_{\leqslant_{ch}}(x) := \{y \in Perd : y \leqslant_{ch} x\}$[2].

9. Notice that I do not assume that $\leqslant_{ch}$ satisfies all the axioms for $\leqslant$. The reason is that characteristic parts are defined by intentional acts performed arbitrarily by members of communities. Thus, $\leqslant_{ch}$ may not satisfy even minimal mereological constraints. On the other hand, $\leqslant_{ch}$ is assumed to be a partial order because reflexivity, symmetry, and transitivity constitute the lexical core of any mereological theory (cf. [4], p. 33-38).

10. There are no mixed genres, i.e. there is no such genre that the set of its supports contains both endurants and perdurants.

    $$Supp \cap End = Supp \vee Supp \cap Perd = Supp. \tag{28}$$

11. The language element of a document genre will be modelled by a function *Lang* that maps a set of sets of equiform endurants into a set of sets of situation-types, i.e. if $X \subseteq \wp(End)$, then $Lang : X \to \wp(Sit)$. This modelling solution is based on four assumptions.

    - We can define on the set of all informational features of documents a relation of equiformity.
    - Any informational feature of any document is endowed with a propositional content.
    - Any such propositional content is built out of propositions.
    - Any proposition functionally corresponds to a situation-type.

    Subsequently, if $X \in Lang(Y)$, then this means that any endurant from Y conveys a piece of information represented by X.

12. The language element of a communication will be modelled by a function *Lang* that maps a set of sets of equiform perdurants into a set of sets of situation-types, i.e. if $X \subseteq \wp(Perd)$, then $Lang : X \to \wp(Sit)$. This modelling solution is based on the same assumptions as in the previous remark.

[2] Although I use the same symbol for the relation of being a characteristic part of a document in the strict sense and the relation of being a characteristic part of a communication, it must be remembered that they are actually two different relations. The same remark applies to the symbol "*Lang*" introduced later on.

13. Although I will not provide any detailed description of *Lang*, let me just observe that modelling genre languages by means of such functions, one need to specify under which conditions two endurants (perdurants) are equiform. Here it suffices to claim that the relation of equiformity is an equivalence relation. This implies 29:

$$X_1, X_2 \in domain(Lang) \rightarrow X_1 \neq \emptyset \wedge X_2 \neq \emptyset \wedge X_1 \cap X_2 = \emptyset. \tag{29}$$

Moreover, any support of any genre should contain at least one informative part:

$$\forall x \in Supp\ \exists y\ [y \leqslant x \wedge y \in \bigcup domain(Lang)]. \tag{30}$$

14. The work of Yates and Orlikowski and the above formal framework assume that any genre is associated with exactly one language. Since this assumption seems too strong, we may treat any function *Lang* as the "sum" of all languages associated with a given language. Obviously, this solution presupposes that we dealt somehow with those word-inscriptions that in different languages convey different meanings (e.g. "was" in English and German).

A *genre* x from a community y is a pair $< Use, Content >$ such that:

1. $Use =< Trigger, Purpose >$, where
   (a) $Trigger \subseteq Sit \wedge Trigger \cap Ment\_Sit \neq \emptyset$,
   (b) $Purpose \subseteq Sit \wedge Purpose \cap Ment\_Sit \neq \emptyset$,
   (c) $Ment\_Sit(y) \cap (Trigger \cup Purpose) \neq \emptyset$,
2. $Content =< Med, Lang >$, where $Med =< Supp, \leqslant_{ch} >$.

Philosophical caveat. For a philosophically conscious reader, I should add that the above definition is to be interpreted as "A genre ... is represented as a pair ...". Strictly speaking, a genre x from a community y represented as above is an intentional entity such that

1. x generically constantly depends in its existence on the beliefs of the members of y,
2. for each trigger z for x, at least one member of y holds a belief that is equivalent to the belief that z is a trigger for x,
3. for each purpose z of x, at least one member of y holds a belief that is equivalent to the belief that z is a trigger of x,
4. for each support z of x, at least one member of y holds a belief that is equivalent to the belief that z has the characteristic parts specified by $\leqslant_{ch}$,
5. at least one member of y is a competent user of the language represented by *Lang*.

Similar remarks apply to all definitions below.

A genre $x =< Use, << Supp, \leqslant_{ch} >, Lang >>$ from a community y is a *document genre* iff $Supp \subseteq End$. A genre x$=< Use, << Supp, \leqslant_{ch} >, Lang >>$ from a community y is a *communication genre* iff $Supp \subseteq Perd$.

x is a *document of a genre* $< Use, << Supp, \leqslant_{ch} >, Lang >>$ iff $x \in Supp$.

Besides the constraints introduced above, I submit four axioms: 32, 33, 34, and 35, in order to exclude communicationally unreasonable cases of genres. Notice that all these axioms refer to genres enacted within a single community.

Any genre is to encompass all documents that share the same structure with respect to their characteristic parts provided that their other genre-related aspects are identical. This condition is equivalent to axiom 32 below. In order to put it in a concise way, I use the following auxiliary definition. Let X be a set of genres from a given community. Let $x_1 =< Use_1, << Y_1, \leqslant_{ch_1} > , Lang_1 >>$ and $x_2 =< Use_2, << Y_2, \leqslant_{ch_2} >, Lang_2 >>$ belong to X.

$$x_1 \approx_X x_2 \equiv \tag{31}$$
$$\equiv \exists y_1 \in Y_1 \exists y_2 \in Y_2 < P_{\leqslant_{ch}}(y_1), \leqslant_{ch_1} > \ is\ isomorphic\ to\ < P_{\leqslant_{ch}}(y_2), \leqslant_{ch_2} > \wedge$$
$$Use_1 = Use_2 \wedge Lang_1 = Lang_2.$$

Notice that the relation defined by 31 is an equivalence relation in X. Let $[x]_{\approx_X}$ be a $\approx_X$-equivalence class containing $x \in X$.

$$\forall x \in X | [x]_{\approx_X} | = 1. \tag{32}$$

The characteristic parts of a given genre are selected in order to mirror the social and informative functions of this genre. For example, the characteristic parts of business letter reflect the cultural relations within a given community and the economic interests of its members. Therefore, if two genres share their use components, then they ought to share their characteristic parts provided that the sets of their supports are identical. Enacting (within a single community) two genres such that they share their use and support components, but which differ in their characteristic parts, would be communicationally ineffective. Let $< Use_1, << Supp_1, \leqslant_{ch_1} >, Lang_1 >>$ and $< Use_2, << Supp_2, \leqslant_{ch_2} >, Lang_2 >>$ be two genres from one community.

$$Use_1 = Use_2 \wedge Supp_1 = Supp_2 \rightarrow \leqslant_{ch_1} = \leqslant_{ch_2} . \tag{33}$$

Conversely, if two genres share their characteristic parts, then they ought to share their use elements. If two genres shared their characteristic parts, but differed in their use components, this would mean that the set of characteristic parts of one of these genres should be extended in order to discriminate the social functions of one of these genres from the social functions of the other. (Remember that by definition that $\leqslant_{ch_1} = \leqslant_{ch_2}$ implies that $Supp_1 = Supp_1$.) Let $< Use_1, << Supp_1, \leqslant_{ch_1} >, Lang_1 >>$ and $< Use_2, << Supp_2, \leqslant_{ch_2} >, Lang_2 >>$ be two genres from one community.

$$\leqslant_{ch_1} = \leqslant_{ch_2} \rightarrow Use_1 = Use_2. \tag{34}$$

Finally, because all supports of a genre are created in order to convey information relevant for the community that enacted this genre, two genres with the same supports sets and use elements should be identical with respect to their languages. Otherwise, it would follow that the community in question may "decode" the same set of documents that it enacted in two different languages even when the community uses these documents in the same circumstances and ascribes the same purposes to them.

$$Supp_1 = Supp_2 \wedge Use_1 = Use_2 \rightarrow Lang_1 = Lang_2. \tag{35}$$

Axioms 32, 33, 34, 35 entail that two genres from one community are identical iff their characteristic parts are identical.

We are now in a position to define several general types of genres.

A document x of a genre $< Use, << Supp, \leqslant_{ch} >, Lang >>$ from a community z is *simple* iff $P_{\leqslant_{ch}}(x)$ is a singleton. A document x of a genre y from a community z is *complex* iff x is not a simple document of y from z.

A genre $<< Trigger, Purpose >, Content >$ from a community x is

- *internal for* x iff $Ment\_Sit(x) \cap Trigger \neq \emptyset$, $Ment\_Sit(x) \cap Purpose \neq \emptyset$, and $\forall y \in Com\ [y \neq x \rightarrow Ment\_Sit(y) \cap (Trigger \cup Purpose) = \emptyset]$,
- *external for* x iff either $Ment\_Sit(x) \cap Trigger = \emptyset$ or $Ment\_Sit(x) \cap Purpose = \emptyset$.

A genre $<< Trigger, Purpose >, Content >$ from a community x is

- *informative for* an agent y iff $\exists z \in Sit < y, z > \in Purpose \cap Bel$,
- *motivational for* an agent y iff $\exists z \in Sit < y, z > \in Purpose \cap Des$,
- *prescriptive for* an agent y iff $\exists z \in Sit < y, z > \in Purpose \cap Int$.

A genre $<< Trigger, Purpose >, Content >$ from a community x is *socially informative for* x iff $\exists y \in Sit\ \forall z\ in\ x < z, y >, < z, < z, y >> \in Purpose \cap Bel$.

A genre $<< Trigger, Purpose >, Content >$ from a community x is *socially expressive* iff there is $y \in Sit$ such that

1. $\forall z\ in\ x\ \ < z, y >, < z, < z, y >> \in Bel$,
2. $\forall z \in Supp\ \exists X\ \exists v \in X\ [v \leqslant z$ and $y \in Lang(X)]$.

A genre $<< Trigger, Purpose >, Content >$ from a community x is a *means of information flow* from x to a community y iff for some $z \in Bel$, $z \in Trigger \cap Ment\_Sit(x)$ and $z \in Purpose \cap Ment\_Sit(y)$.

Let X be a non-empty finite set of genres, i.e.
$X := \{<< Trigger_i, Purpose_i >, Content_i >\}$, for $i \in \mathbb{N} \setminus \{0\}$. I will say that X *constitutes a genre system* iff for any j, $1 \leq j < i$, $Purpose_j \cap Trigger_{j+1} \neq \emptyset$.

A *genre repertoire* of a community x is a pair $< Comp, Freq >$ such that

1. $Comp$ is the set of all genres from x,
2. $Freq$ is a function $Freq : Comp \rightarrow \mathbb{Q}$ such that
   if $y =< Use, < Supp, \leqslant_{ch} >, Lang > \in Comp$, then $Freq(y) := \frac{|Supp|}{|Supp(Comp)|}$, where $Supp(Comp) := \bigcup\{Supp :< Use, << Supp, \leqslant_{ch} >, Lang > \in Comp\}$.

The genre repertoire of a community is determined by the set of all genres enacted in this community and the frequency function that specifies for each genre from this set the proportion of the number of documents of this genre to the number of all documents created within this community.

Finally, we are also in a position to define the relation of genre subsumption. In contradistinction to our previous definitions of the notions used by Yates and Orlikowski, we are now left with no clue as to what it exactly means that one genre subsumes another. Thus, the following definition is highly stipulative.

A genre $< Use_1, << Supp_1, \leqslant_{ch_1} >, Lang_1 >>$ from a community x *subsumes* a genre $< Use_2, << Supp_2, \leqslant_{ch_2} >, Lang_2 >>$ from x (in a social context of x) iff $\leqslant_{ch_1} \subseteq \leqslant_{ch_2}$.

The definition presupposes that the social context to which the relation of subsumption is to be relativised is given by the community parameter. This implies that only genres from the same community can be compared with respect to the subsumption relation.

It is easy two observe that the relation of subsumption is a partial order on the set of all genres from a given community.

5 Conclusions

Searching for the ontological commitments of the theory of genres propounded by J. Yates and W. Orlikowski, I arrived at a formal ontology of genres expressed in set-theoretical terms. Within this ontology, I showed how to represent those aspects of genres and genre documents that were mentioned by Yates and Orlikowski and introduced a few new features. It turned out that the resulting conceptual structure is complex enough to allow us to describe a broad range of communicational phenomena. Nonetheless, the set of ontological categories to which I had to resort is not sparse. The question whether we could describe the same range of phenomena on the same level of precision without such ontologically demanding categories as situation-types and mental attitudes remains open.

References

[1] J. Barwise and J. Perry. *Situations and Attitudes*. A Bradford Book. The MIT Press, Cambridge (MA), 1983.

[2] C. Berkenkotter and T. N. Huckin. *Genre Knowledge in Disciplinary Communication*. Lawrance Erlbaum Associates, Hillsdale (NJ), 1995.

[3] T. Bittner, M. Donnelly, and B. Smith. Individuals, universals, collections: On the foundational relations of ontology. In A. C. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems*, pages 37–59, Amsterdam, 2004. IOS Press.

[4] R. Casati and A. C. Varzi. *Parts and Places*. The MIT Press, London, 1999.

[5] K. Crowston and B. H. Kwasnik. A framework for creating a facetted classification for genres: Addressing the issues of multidimensionality. In *Proceedings of the 37th Annual Hawwaii International Conference on System Sciences*, volume IV, 2004.

[6] K. Crowston and M. Williams. Reproduced and emergent genres of communication on the world-wide web. In *Proceedings of the 32nd Annual Hawwaii International Conference on System Sciences*, volume VI, pages 30–39, Los Alamitos (CA), 1997. IEEE Computer Society Press.

[7] A. Gangemi, S. Borgo, C. Catenacci, and J. Lehmann. Task taxonomies for knowledge content. deliverable D07, Laboratory for Applied Ontology, ISTC-CNR (Italy), 2005.

[8] A. Giddens. *The consitution of society*. University of California Press, Berkeley, 1984.

[9] R. Glushko. *Document Engineering : Analyzing and Designing Documents for Business Informatics and Web Services*. MIT Press, 2005.

[10] B. Heikkinen. Generalization of document structures and document assembly. Technical report, Department of Computer Science, University of Helsiniki (Finland), 2000.

[11] H.-G. Im, J. Yates, and W. J. Orlikowski. Temporal coordination through communication: using genres in a virtual start-up communication. *Information, Technology and People*, 18(2):89–119, 2005.

[12] Roman Ingarden. *Der Streit um die Existenz der Welt*, volume 2. Max Niemeyer, Tubingen, 1965.

[13] K. Jokinen, J. Borgman, and R. Sulonen. Common data model for design document exchange in bussiness-to-bussiness networks. In *Proceedings of the 38th Hawaii International Conference on System Sciences - 2005*, pages 1–1–, 2005.

[14] A. Kennedy and M. Shepard. Cybergenre: Automatic identification of home pages on the web. *Journal of Web Engineering*, 3(3-4):236–251, 2004.

[15] H. Kilov and L. Cuthbert. A model for document management. *Computer communications*, 16(6):408–417, 1995.

[16] R. Klischewski. Towards an ontology for e-document management in public administration - the case of schleswig-holstein. In *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*. IEEE Computer Society, 2003.

[17] D. Lewis. *Convention*. Blackwell Publishing, 2002.

[18] S. Y. Lim, S. B. Park, and S. J. Lee. Document retrieval using semantic relation in domain ontology. In P. Szczepaniak, editor, *AWIC 2005*, LNAI 3528, pages 266–271, Berlin, 2005. Springer-Verlag.

[19] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonderweb deliverabled18, 2003.

[20] C. R. Miller. Genre as social action. *Quarterly Journal of Speech*, 70:151–167, 1984.

[21] W. J. Orlikowski and J. Yates. Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39(4):541–574, 1994.

[22] T. Ryan, R. H. G. Field, and L. Olfman. Homepage genre dimensionality. In *Proceedings of the Eighth American Conference on Information Systems*, pages 1116–1128, 2002.

[23] P. M. Simons and Ch. W. Dement. Aspects of the mereology of artifacts. In R. Poli and P. Simons, editors, *Formal Ontology*, pages 255–276. Kluwer Academic Publishers, Dordrecht, 1995.

[24] J. Swales. *Genre Analysis*. Cambridge University Press, Cambridge, 1990.

[25] R. Tuomela. *The Importance of Us*. Stanford Series in Philosophy. Stanford University Press, Stanford (CA), 1995.

[26] Ch. Welty and J. Jenkins. Fromal ontology for subject. *Data and Knowledge Engineering*, 31:155–181, 1999.

[27] M. Woolridge. *Reasoning about Rational Agents*. The MIT Press, Cambridge (MA), 2000.

[28] J. Yates and W. J. Orlikowski. Genres of organizational communication. *Academy of Management Review*, 17(2):299–326, 1992.

[29] J. Yates and W. J. Orlikowski. Explicit and implicit structuring in genres in electronic communication: Reinforcement and change in social interaction. *Organization Science*, 10(1):83–103, 1999.

[30] J. Yates, W. J. Orlikowski, and J. Rennecker. Collaborative genres for collaboration. In *Proceedings of the Thirtieth Annual Hawwaii International Conference on System Sciences*, volume VI, pages 50–59, Los Alamitos (CA), 1997. IEEE Computer Society Press.