Geometric Theory of Machine Awareness for Legal Information Retrieval and Reasoning

M.A. Heather
School of Law
University of Northumbria at Newcastle NE1 8ST, UK

B.N. Rossiter
Computing Science
Newcastle University NE1 7RU
B.N.Rossiter@newcastle.ac.uk

July 1995

1 Abstract

This report[1] considers that links in hypertext are representable as links in thought by covariant arrows between categories. Taken in dynamic context, the right–exactness of the Heyting implication $\mathbf{A} \Rightarrow \mathbf{B}$ corresponds to inference and the next document in a non–linear trail through hypermedia. Awareness is provided by the dual contravariant arrows with the important special case of the intension–extension relationship. The corresponding left–exactness is the closure limit that invokes consciousness.

About the author

Michael Heather is senior lecturer in law where he has been responsible for computers and law since 1979.

Nick Rossiter is lecturer in the Department of Computing Science with particular interests in databases and systems analysis.

Suggested Keywords

hypertext, legal reasoning, information retrieval, category theory, adjoints, Heyting algebra.

---

[1] The work on geometric logic and law in this report was presented at the 17th IVR World Congress June 16th–21st 1995 *Challenges to Law at the End of the Twentieth Century* held at University of Bologna, Italy in working group 66 on legal computer science. Aspects dealing with legal hypertext have been accepted for publication in the journal *Informatica e diritto*. The contribution that machine awareness can make was given at the post-conference paper for *Consciousness : the Big Issue* at the British Computer Society (BCS), Cybernetic Machine Special Interest Group (SIG) Meeting, held at University of Greenwich. 11th March 1995. and included in the postproceedings.

2 Introduction

The wealth of information available today online cries out for systems to have some awareness and alerting capability to identify automatically relevant documents. The need for an alerting function was recognized early in law [Bing & Harvold 1977] where it has always been necessary to handle large quantities of data. This is now a problem which faces an average user of electronic mail.

Current tools to explore and handle information overload on the InterNet (like GOPHER, WAIS, USENET, MOSAIC, TELNET and ARCHIE) depend on preset buttons, predetermined indexes and algorithms, statistical clustering or human intervention. These methods are proving inadequate. Systems now need to exhibit some characteristic of self–consciousness. They need to be aware of their own contents to provide the user with dynamic links to be made in context at run–time. The current state of the art [Nielsen 1990, Manager 1995] like the use of dynamic frames has more to do with formatting displays than connecting and identifying relevant content. The hypertext markup language HTML of World–Wide Web provides the facility of tags for connecting to related material elsewhere on the Web but the user has to provide the means for identifying the material. Enhancements of SGML and proprietary software like HEADS from Hitachi, DotBook from Novell, HotJava from Sun, and Cairo in the new version of Windows NT [Rymer, Guttman & Matthews 1994], are approaching the concept of mark–up in a semantic context. New environments and operating systems like Windows95 come with built–in facilities for network surfing.

However, users need to be assured of the quality of their legal information system. The quality controller like in any industrial process has to be at a separate supervisory level. This is a trigger mechanism in the system to identify relevant information in context and is also a self awareness where the information checks itself for completeness and its own limitations. Intelligent hypertext is an initial step at this level but intelligence is insufficient without a layer of consciousness. The quality assurance level is a closure over all participating sub-systems, whether local to the end-user or global and belonging to the information providers.

The information in the subsystems may come in any form or format. Of great importance for law is the image data found in multimedia for the large quantities of documents that are being input by scanners in legal cases [Chepalis 1994]. A hypertext system that cannot search, identify and retrieve the contents of documents held as image bits has little value for lawyers. A move in this direction to provide features for heterogeneous data can be seen in systems like HyperNet [Marovac & Osburn 1992].

Without proper procedures, the *less paper office* can result in a loss of integrity. To ensure consistency, some formal model is needed to underpin the interoperable subsystems [Heather & Rossiter 1993]. As information systems are real–world and open, the models need to be drawn from constructive mathematics [Heather & Rossiter

1994] where intuitionistic logic seems to be able to give a high formalism to common sense reasoning and experience.

As legal information systems become larger and more complicated it is very easy for the user to get lost in them. Even intelligent hypertext becomes inadequate. It can provide the user with a non-linear connectivity but to be of value the system needs to know where the user is, where the user has come from and where the user is going, all relative to the contents of the information at any point.

3 Types of Awareness

But the system also needs to provide a variety of awareness features. There is the nature of the information relative to the informational needs of the user and incidentals like the appropriate methods of displaying the data relative to the users individual preferences. More important is the self-awareness of the information relative to information elsewhere, for example to a source of continuous updates. Otherwise, hypertext can misinform.

Whether computers will in the future exhibit the characteristics of human consciousness is an interesting subject of speculation. Of more current importance is the need for information systems to have a current awareness of the contextivity of the information that is available. This encompasses both local disks, CD ROM, etc under the user's physical control as well as the interaction of these with any central online facilities over which the user has electronic control at least as far as access and availability.

Even with the old printed medium there were and still are various levels of information providers and information users. At one end, there is the consciousness that a book seller has of the availability of information extending little beyond a stock of items currently in market demand or potential demand together with a list of titles and authors that can be ordered. Unless specialized, the bookseller's awareness will not normally extend beyond books in print. Librarians have more regard for the content of the information but this information may not go far beyond books currently in print and the use of bibliographies.

A reference librarian on the other hand has more interaction with the contents of the information. That is the librarian actually opens the books in question. At the extreme is the lawyer whose function is to dispense the information in the form of advice involving the construction and application of legal source documents to clients who may never actually see the books themselves or even know that they exist. The lawyer has to have an appreciation of knowing what exists and how to find it. Because of the volume and complexity of the information available, the awareness carried around in the lawyer's head is only how to go about finding some relevant piece of information and not any particular source. It is this sophisticated aim to make conscious reasoned connections that legal hypertext must address.

3.1 Conscious Reasoning as Connections

Legal reasoning lies at the heart of the practice of Law implicit in routine procedures and explicit in activities like legal research, legal information retrieval and legal decision making.

Each age has its own view of logic and the application of logic in the wider concept of reasoning. Reasoning today is making connections: making connections between different categories of objects which may be mixed abstract and concrete. To reason is to impose a structure on the objects. This order may be one of many possible. There may be a unique path but in the law there is usually not just one way to reach the same conclusion and sometimes there may be equally valid paths to other conclusions. In terms of discrete mathematics, the ordering is to identify relationships between objects.

Whether the objects and relationships are considered separately or wrapped-up together is a matter of viewpoint. The Anglo-Saxon word *thoughts* expresses this dichotomy perhaps even better than the word *reasoning*. Comparable perceptions are to be found in other modern relevant topics such as the object-oriented paradigm and connectionist models like parallelism and neural nets.

The older concept of the continuum has been displaced by discrete objects partly because of strong digital influence from computer science and partly arising from a better understanding of physical processes. A critical feature in this reasoning is the connections that pass through real-world events. This provides a source of variety and normally determines that no particular path is fixed for all time.

3.2 Hypertext Connections need to reflect Real–world Links

Inferences in making the right connections involve identifying the objects to connect. These are heavily dependent on pragmatic considerations which may be very subject to the peculiarities of the situation or there may be a more generalizable context sensitivity. Take a question involving law and economics.

Who pays a higher rent: tenants of controlled or uncontrolled tenancies?

Controlled tenancies are connected to security of tenure so that tenants subject to a controlled tenancy have security of tenure whereas the uncontrolled do not have the same security. As security is more valuable, the first impression might be that controlled tenants will pay more for the tenancy if secured. In practice, secure tenants on average pay much less than unsecured tenants. It is a matter of making the right connections. For controlled tenants also have their rents regulated so that they cannot be increased. The real-world situations that need to be connected are: because of the demand for housing, those who cannot obtain controlled tenancies have to take tenancies where there is no control over the rent charged.

This is a kind of context sensitivity that shapes the result that uncontrolled tenants pay more. This could be determined by reference to statistical data of mean rents for the different types of regulated property, However, a few moments reflection will show that no empirical knowledge is needed. The correct conclusions can be made by making the right connections. Again, these connections are the only ones that could be made, namely that controlled tenancies only exist in times of shortage for otherwise market forces will keep the rents low.

What form of mapping do the relationships take? Essentially it would seem that they depend on some kind of pattern matching. At the simpler level, it is the identification as in the sequence of characters in connections made through vocabulary. Higher order connections can be made at a very abstract level as in a pattern of legal objects and events or facts which by induction make a new legal object and which can by a connection of transference amount to a reasoning by analogy. It is the basis of hyperspace navigation that the low-level connections and the high-level reasoning processes are essentially equivalent classes. The question of the practical technology is how to handle these objects and their identities. This is a class of problem relevant to any implementable information system, including knowledge bases, deductive databases, advanced expert systems, as well as hypertext.

The connections are made in a coordinate space which is not cartesian and where the points are objects of some complexity. A human being is a good example of a complex object consisting of a bundle of characteristics with a unique identification. External positions change with time. Internal characteristics depend on time and also on external positions.

4 Law as Structure of Identifiable Objects and Relationships

4.1 Objects and Subobjects

What are the legal objects? They appear to be recursive structures that can be analyzed in various ways as, in common law jurisdictions, principles, norms, concepts, authority, precedent, *ratio decidendi, obiter dicta*, primary and secondary rules, duties, obligations, responsibilities, rights. There are finer distinctions, for instance between the fiduciary duty of a trustee and that of an agent. Common law cases are very full instances of legal objects forming nodes in a network approximating to a Boolean lattice of citations or some more complicated partially-ordered set of a non-linear chain of inference. While the document is an object, it is also composed of many interesting concepts that may be legal objects in their own right. The document is an aggregation of these subobjects where it has to be noted that the position of a document in the network normally depends on the connections between the subobjects. A single document may be a member of quite different networks because of quite disparate subobjects.

The primary sources of the common law of England at present consist of highly–structured documents, cases and statutes inter–locked together [Rossiter, Sillitoe & Heather 1990] by the subobjects they contain. These are the principles, concepts and arguments that have to be connected to arrive at the conclusion to resolve a legal problem. In order to represent readily the complex relations, both statutes and case reports employ an elaborate presentation of text with diverse formats and multiple character sets and sizes. To capture adequately the detailed structure of the relationships embedded in this text, sophisticated modelling techniques are required to retain the information and transform it to a machine representation.

Database techniques can provide in electronic publishing the typing corresponding to typography. The printed version presents a fixed view of the structure which has to be accepted in a common form by all readers although different views will be in their minds as they read the text. The organization of the text data as appropriate for different users, however, is now available by exploiting the power of computers for logical re-organization. By holding the complete structure of each statute and case in database records, for example, the display format can be readily changed to meet new circumstances: presentation of output to the human eye requires quite different forms from operation in a machine-machine mode, where the output is directly into a word-processor or mark-up language [Heather 1989], or where one machine interrogates another for information [Connolly 1985].

4.2 Connections in the Law

Hyperspace is equivalent to a multimedia database composed of the complex objects described above in contrast to traditional types of data. Databases for simple data developed out of advanced file handling and it was soon realised that it was necessary to identify the kind of relationship which existed between data. The traditional structures are the network, hierarchical and the relational model of Codd. With networks based on the theory of directed graphs, the hierarchical or nested trees and the relational model relying on mathematical sets, these are still only lean representations of relationships in the real world.

Semantic models have developed to meet the need to specify information about the kind of relationships like the way that the relational model was extended in Codd's Tasmanian model RM/T [Codd 1979] to capture more meaning through the introduction of rules of relationships and integrity. The most popular of the semantic models is probably the E-R model of Chen but this has been extended in different ways to suit different users [Spaccapietra 1987]. For instance the type and attribute form had to be added to Chen's original style of representation and Sakai added generalization. The important deficiency of manipulation and the specification of behavioural characteristics has been satisfied in later models by Taxis, Event and SHM+. The complex nature of the objects on the other hand has been satisfied by the development of object-oriented models such as the Semantic Association Model SAM* for use in statistical databases.

They show the way and some of these models have already been applied to document structures for example E–R [Heather & Rossiter 1987], [Rossiter & Heather 1988], extended relational and Taxis [Rossiter & Heather 1989], and object–oriented and SQL relational [Rossiter & Heather 1990]. None of the models have all the necessary features. The kinds of relationships needed are:

- abstractions such as aggregation, generalization, specialization, inheritance, classification, definition, designation, associative;
- structural such as models, nets, tables, hierarchies, entity classes, E-relations, P-relations;
- statistical such as summation, averages, probabilistic, fuzzy;
- ordering such as sequence, Markov chain, probabilistic, temporal, stochastic;
- reductionist such as projection, parallax, derivation, view;
- behavioural such as dynamic, functional, transaction, operational;
- synthetic such as composition, join, union, cross-product, combined, concatenation, insertion, injection, embedded, tributory;
- analytic such as selection, intersection, adjacency, parametric, attribution;
- parallel such as synchronization, collaborative, collateral, adjacency, adjoin, redundant, orthogonal, anti-parallel, contributory.

Database technology has made progress with some but not all of the above categories. In some relations, the user is not concerned with detailed procedures and these may be the beginnings of automated reasoning. For instance, in aggregation the user is unaware of the way in which the subobjects are put together. An example of this is the way that the current version of a section of an English Act of Parliament may be derived from a number of textual amendments in later Acts. Automatic identification of objects and their characteristics is needed to make the selection with the right interconnections for the aggregation. From a database point of view, the identification is provided by the keys in the system [Rossiter & Heather 1988]. Some universally recognized form is necessary to recognize the keys. This enables documents to be addressed and cross–referenced in a natural manner with a standard identification mechanism. Data typing can be used to characterize components of the keys so that documents can be composed from their underlying subsections (subobjects) in a transparent way. The use of natural keys and relations avoids the unnecessarily reductionist methods of early legal retrieval systems.

Elaborate data management systems are needed to provide the high functionality required for structuring, manipulating and maintaining the data with the necessary integrity to provide professional information systems so that end-users may have access transparently to goal-information in a highly organized state. To do this the management system has therefore to recognize inherent relationships in the data to make the necessary hypertext links.

There are often very many, if not an infinite number of, natural connections that can be made. The author or information provider may predefine certain of these based on some expectation of the user's requirements. Alternatively it may be possible to provide some automatic assistance based on predetermined criteria. It is a simple matter to have a dynamic button to pick up references for a glossary or thesaurus where there is a direct connection usually because the item in the text is itself a simple key to the citation. Where there is a partial or a composite key, the system has to have some awareness functionality of what is needed.

Both these methods are limited. The system needs to be able to follow any potential connection under the control of the user. This requires the system to be conscious of where the user is within the document. A very simple example might be given of anaphora when reading from a Hansard CD-ROM, to deal with a sentence beginning with

"As I said in my speech on 28th October to this House ..."

the system needs to be aware of the name of the speaker, of whether the current speech is being made in the House of Commons or the House of Lords and of the date of the speech to identify the appropriate year for the 28th October. This necessary awareness required is therefore beyond intelligent hypertext. It also illustrates the practical point that this awareness needs to be a runtime facility. For identifying all possible cross references in advance when only very few of them will ever be required is very inefficient in preprocessing and storage and almost impossible manually.

The law is typical of a number of areas of information where there is a large store of static data which can be kept locally cheaply on video-disk or CD-ROM but where it is essential to be able to check the present state against some master version. Take the example the legal question of defining an infringing copy of a piece of software. The current law is to be found in the Copyright, Designs, and Patents Act 1988. When it came into force it confirmed software as having the protection as that of a literary work. Section 27 defined the meaning of 'infringing copy' without any special reference to software.

However under the Copyright (Computer Programs) Regulations 1992 (Statutory Instrument 1992 No 3233) which came into force on 1st January 1993 the 1988 Act was amended to vary the definition of 'infringing copy' from the meaning for other literary works to bring English law in line with the European Directive No.91/250/EEC (O.J. No L122, 17.5.91, page 42). This situation shows that some awareness and alerting function is necessary to know that some later version is needed. This is only one way in which amendments are prepared in English law. All of which need to be recognizable by the system.

This awareness is now essential in very many areas of business which need continual access to information on changing standards and regulations. This awareness function can be achieved by overlaying another layer of metadata on top of the basic hypertext system. This is a necessary part of intelligence in information retrieval systems [Heather 1985]. We have to provide this additional layer to simulate a human metamemory for any type of document [Rossiter 1987] and to be reliable as comprehensive as one that like consciousness provides a closure to an open system [Heather 1988].

5 Formal Modelling under Geometric Logic

The logical reasoning obtained with axiomatic methods are subject to the uncertainties of the applicability of the axioms. Constructive mathematics on the other hand attempts to develop logically what works in practice and can perhaps provide the necessary universality for interoperability of heterogeneous data systems with consistency and quality assurance in the real–world. Geometric logic is particularly appropriate for modelling relationships in hyperspace [Heather & Rossiter 1994] for it is essentially concerned with links between objects.

From the simple concept of the arrow, formal categories can be constructed of objects with arrow links between them. These provide a natural model for a document. Geometric logic is the formulation between the categories and can therefore represent manipulation of documents in this model of hyperspace. In this model it turns out that linking documents and reasoning are equivalent. This confirms our earlier informal discussion.

We are concerned with general categories which may be used to represent any system or a class, object, entity, set, etc. that satisfy the four categorical axiomatic constructs for arrows namely composition, compatibility, associativity and compositional identity [Barr & Wells 1990; Heather & Rossiter 1994]. These required constructs do not cause many problems in applying category theory to real–world models which deal with things that actually exist. However, it may be necessary to check carefully that the components of a virtual reality system satisfy the definitions of a category. Because categories are general it is often only a matter of convenience for a particular model how objects and arrows are to be identified. With hypertext a document forms a natural category. Other categories are always available to provide the necessary typing. For example a particular Act of Parliament can be considered as a category **Act**. **Act** can be typed by an arrow from the category **Enact** of statutes consisting of all parliamentary enactments including both civil law codes (from the subcategory **Code**) as well as statutes (of the subcategory **Stat**) found in common law jurisdictions.

Figure 1 shows functor arrows K, L between categories $\mathbf{A}$ and $\mathbf{C}$ containing objects $A, B, C, \ldots$ interrelated by arrows $f, g, \ldots$. In Figure 1, K assigns from the source object A the target object K(A) to C and from a source arrow f the target arrow

K(f) to g. These are covariant arrows. The direction of K and L may be reversed to give the dual contravariant arrow.

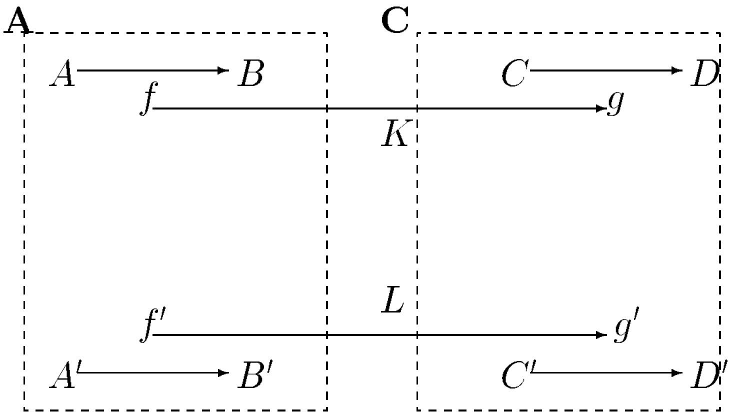

Figure 1: Functors compare Categories

Documents and the concepts they contain may be represented by categories $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$. The functor arrow generally represents a hypertext link between documents. The functors can also represent inferences. In geometric logic a deductive system is based on the arrow as a proof:

$$\frac{F : \mathbf{A} \longrightarrow \mathbf{B} \qquad G : \mathbf{B} \longrightarrow \mathbf{C}}{GF : \mathbf{A} \longrightarrow \mathbf{C}}$$

where $F : \mathbf{A} \longrightarrow \mathbf{B}$ is more than a proof theory entailment. The arrow is the *reason* why $\mathbf{A}$ entails $\mathbf{B}$ [Lambek & Scott 1986]. The *reasons* F and G may be a mixture of propositional and predicate logic or even deontic. The inference in the composition GF is then a graph in geometric logic of legal reasoning. At the same time it is calculable through the algebraic form.

Any model needs a philosophical view to be taken of the universe. Hypertext too cannot exist in a philosophical vacuum. Therefore it is perhaps worth mentioning in passing the ontology of this formal model. Reality in the everyday world is made up of rational links as discussed in the first part of this paper. What exists are limits in the sense of geometric logic. A hypermedia is a model of these limits in some cyberspace. These can be represented in this way because of the universal abstract character of category theory.

In general a finite limit in the category $\mathbf{C}$ means a limit of a functor $\mathbf{J} \longrightarrow \mathbf{C}$ where $\mathbf{J}$ is a finite category. An object in the functor category $\mathbf{C}^{\mathbf{J}}$ is a geometric diagram in $\mathbf{C}$ of type $\mathbf{J}$ which can be represented in general by the cone (together with dual cocone) [Rhydeheard & Burstall 1988].

The nature of proof in category theory should be emphasised. The diagram is a formal diagram. It is a geometric representation equivalent to an expression in algebra. We are in constructive mathematics and the one proof needed is the proof of existence. Therefore so long as it can be shown that the entities belong to formal categories [Freyd 1964], proof up to natural isomorphism is by composition. A formal diagram is in effect a *quod erat demonstrandum.* Freyd & Scedrov [1990 p.29–36] give a formalization of the diagrammatic language.

5.1 Adjointness

Adjointness between two categories

$$F \dashv U : \mathbf{A} \longrightarrow \mathbf{B}$$

has left and right components which specify how an arrow in category $\mathbf{A}$ is related to an arrow in category $\mathbf{B}$. This is the fundamental concept of implication to be found in geometric logic. The left component is the free functor $F : \mathbf{A} \longrightarrow \mathbf{B}$ and the right component the underlying functor $U : \mathbf{B} \longrightarrow \mathbf{A}$. F is left adjoint to U and U is right adjoint to F. This is a natural bijection between arrows which holds subject to the condition for all objects $A \in \mathbf{A}$ and all $B \in \mathbf{B}$ such that:

$$F(A) \longrightarrow B \text{ implies and is implied by } A \longrightarrow U(B)$$

Written as a geometric logic inference where the double line indicates the biconditional (iff):

$$\frac{1_{\mathbf{A}} \leq UF}{\overline{FU \leq 1_{\mathbf{B}}}}$$

With this condition there are two natural transformations or unit of adjunction:

$$\eta : 1_{\mathbf{A}} \longrightarrow UF, \quad \epsilon : FU \longrightarrow 1_{\mathbf{B}}$$

Adjointness is particularly relevant to hypertext for it represents the concept of relative ordering which is the basis of the connections between documents. $A \leq B$ means B is a later document than A in a hypertext trail. The unit of adjunction is a natural transformation that amounts to an abstraction of the components of the adjointness representing the concepts, objects, message passing, etc which connect the documents.

By virtue of the adjoint functor theorem [Freyd & Scedrov 1990], left adjoints preserve colimits (right–exactness) and right adjoints preserve limits (left–exactness). Colimits are the dual of limits. Both limits and colimits will be examined in more detail.

A basic form of awareness can be provided by means of indexing. The simplest index is an inverted file (concordance) which is an example of adjointness, $F \dashv U : \mathbf{A} \longrightarrow \mathbf{B}$. If category $\mathbf{A}$ is a collection of documents where the objects are words, the arrows would be the relationships of natural language with higher–order arrows relating documents. The free functor F is the (arbitrary) addressing for the position of each word in the full text in the documents as adopted by early systems in law like LEXIS. The objects of category $\mathbf{B}$ could also be words and addresses (logical or physical disk addresses); the arrows of $\mathbf{B}$ would be to provide some ordering, usually lexical. The underlying functor U finds for any given word its position in the text of a document as shown in Figure 2:

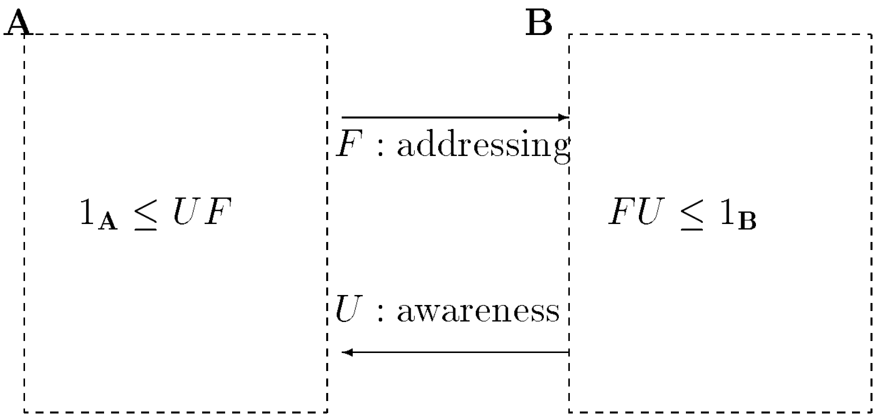

Figure 2: Adjointness in Indexing

Adjointness is satisfied because a given text may be completely decomposed into its vocabulary (right–exactness under functor F) and then reconstituted again in natural language sequence (left–exactness under functor U); that is $1_{\mathbf{A}} \leq UF$. Alternatively starting from the index, the full text can be constituted and then decomposed again; that is $FU \leq 1_{\mathbf{B}}$. Notice that $1_{\mathbf{A}} \leq UF$ consists of all the orderings in the text and $FU \leq 1_{\mathbf{B}}$ all the orderings in the index.

Therefore the contravariant functor $U : \mathbf{B} \longrightarrow \mathbf{A}$ provides the overall awareness of the contents of the documents in category $\mathbf{A}$. LEXIS like most full–text systems has a stop–list of common words to save index storage and therefore operates on a subcategory of $\mathbf{A}$ and also neglects arrows representing other relationships between and within documents. The awareness of these can be retained with a more elaborate database management model [Heather & Rossiter 1987]. Consciousness goes a stage further to identify relevant contents.

5.1.1 Adjointness between Text and Image Data

Imaging is rapidly becoming a major industry and the manipulation of image data based on content and meaning is a burning research topic. Geometric logic shows well the adjointness between textual and graphical information. Both are mapped into the electronic medium as a bit stream.

$$\begin{tikzcd} & \mathbf{GRF}(\mathbf{D}) \arrow[dl] \arrow[dr] & \\ \mathbf{TXT}(\mathbf{X}) \arrow[rr] & & \mathbf{E}(\mathbf{2}) \end{tikzcd}$$

Figure 3: Adjointness of Electronic Forms

Multimedia are logical rather than physical based. They are therefore an abstract category of a document which may be represented as a textual file or as an image file resulting from input by means of a scanner. Clearly the two forms contain equivalent information although they would appear in quite different electronic forms. This is an important example of adjointness as demonstrated in Figure 3. $\mathbf{TXT}(\mathbf{X}), \mathbf{GRF}(\mathbf{D})$ and $\mathbf{E}(\mathbf{2})$ are categories corresponding respectively to text, graphics and electronic form. Each of these categories is a free functor. $\mathbf{TXT}(\mathbf{X})$ is a map from the alphabet X on to finite strings so a character, x, goes to a string, $x \mapsto < x >$. $\mathbf{E}(\mathbf{2})$ is correspondingly composed of strings of zeros and ones. $\mathbf{GRF}(\mathbf{D})$ is the much more interesting graphical version which contains all the semiotic significance of the text beyond the mere characters (i.e. punctuation, capitalization, italics). There may be a loss of information from the category $\mathbf{GRF}(\mathbf{D})$ to $\mathbf{TXT}(\mathbf{X})$.

5.2 Intension–Extension Mapping

The links in hypertext may be at different levels. The mappings representing the links would therefore need to be typed in geometric logic. There is the simple linking between documents like a citation of a label or name (the intension). A more powerful level of connection is within the semantics (the extension). There is also the intension–extension relationship which has been shown by Lawvere [1969] to be composed of adjoint contravariant functors.

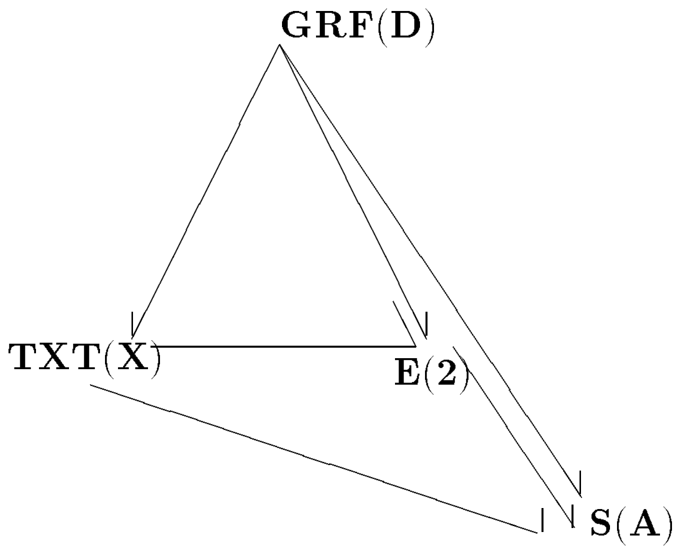

Figure 4: Adjointness in Real–world Semantics

The extension level of the abstract document is therefore the same for the three categories of text, graphics and electronic bits. Equality in geometric logic is provided for by composition. The possible relationships between the three categories of documents at the two levels can therefore all be summed up in a simple geometric formal diagram.

A real–world semantics $\mathbf{S}(\mathbf{A})$ can be represented in any of the three forms of graphical, textual and electronic. There will therefore be intension, and extension consisting of contravariant functors between each of the three and $\mathbf{S}(\mathbf{A})$ as in the diagram in Figure 4.

5.3 Geometric Database Models

Database modelling reduces to a small family of concepts in geometric logic. The various types of database relations described above may be summed up in Table 1. For fuller details that have been worked through for a product model based on limits, see [Rossiter, Nelson & Heather 1994; Nelson & Rossiter 1995].

database operation	categorical construct
abstractions	exactness
structural	adjointness
statistical	subobject classifier
ordering	adjoint functors
reductionist	co–exactness
behavioural	comma category
synthetic	exactness
analytic	co–exactness
parallel	adjointness

Table 1: Database Concepts in Categorical Terms

6 Formal Contextual Sensitivity

6.1 Limits, Colimits and Context

A very fundamental concept that has only been appreciated in the last thirty years is that of *limits* and *colimits* [Mac Lane 1971]. In arithmetic a limit is constructed by multiplication and colimits by addition, Within set theory, intersection is an example of a limit and disjoint union a colimit. With more general categories, limits and colimits become very powerful. A colimit is a deconstruction and provides no new information. The colimit of $\mathbf{A}$ and $\mathbf{B}$ is given by the fullest possible combination of taking them together and written $\mathbf{A} + \mathbf{B}$. A partial colimit would be obtained by taking together only certain parts of $\mathbf{A}$ and $\mathbf{B}$. The parts that are significant when taken together may be provided by the context of a different category $\mathbf{C}$. The pushout $\mathbf{A} +_{\mathbf{C}} \mathbf{B}$ as shown in Figure 5 then expresses this colimit in context.

$$
\begin{tikzcd}
 & \mathbf{A} \arrow[dr, "\iota_l"] & \\
\mathbf{C} \arrow[ur, "K"] \arrow[dr, "L"'] & & \mathbf{A} +_{\mathbf{C}} \mathbf{B} \\
 & \mathbf{B} \arrow[ur, "\iota_r"'] &
\end{tikzcd}
$$

Figure 5: Diagram of Pushout of **A** and **B** over **C**

This is the geometric logic representation of the hypertext link which brings together the documents **A** and **B** through the context **C**. Note that this does not give any new information, but only identifies those parts of **A** and **B** which are relevant together in the context of **C**.

An example of pushouts can be seen in the geometric logic representation of the remark referred to earlier: "As I said in my speech on 28th October to this House ...". The diagram in Figure 6 shows a pasting together of pushouts in which the result of one pushout *House* is included in turn in another pushout forming I.

$$
\begin{tikzcd}
\mathit{year} \arrow[r] \arrow[d] & \begin{array}{l}\mathit{Hansard}\\ \mathit{Record}\end{array} \arrow[r] \arrow[d] & \mathit{Name} \arrow[d] \\
\mathit{28\ \ Oct} \arrow[r] & \mathit{House} \arrow[r] & I
\end{tikzcd}
$$

Figure 6: Composition of Pushouts for
"As I said in my speech on 28th October to this House ..."

New information attained by linking **A** and **B** is given by the product limit $\mathbf{A}\times\mathbf{B}$. This for a context **C** is the pullback $\mathbf{A} \times_{\mathbf{C}} \mathbf{B}$ shown in Figure 7. In general the difference between a limit and colimit may be summed up in that a limit produces some creative outcome of a link whereas the colimit is a link between standard information.

$$
\begin{tikzcd}
 & \mathbf{A} \arrow[dr, "K"] & \\
\mathbf{A} \times_{\mathbf{C}} \mathbf{B} \arrow[ur, "\pi_l"] \arrow[dr, "\pi_r"'] & & \mathbf{C} \\
 & \mathbf{B} \arrow[ur, "L"'] &
\end{tikzcd}
$$

Figure 7: Diagram of Pullback of **A** and **B** over **C**

Hypertext is a family of trails and it is important to recognize whether two trails are distinct or whether they merge. Thus there may be two parallel links between the same two documents. The question arises for the hypertext system whether two separate trails arriving at the first document are then merged. For example a legal case may cite a second case more than once during the report but it may be on two quite distinct points of law or even branches of law. Two cases may be connected on a substantive point of law and also quite separately on a point of legal procedure, adjectival law. A document often cites another more than once. Links between two documents in this situation become a limit point in the two trails if they merge there. However, geometric logic shows that there is a duality of limit also in this instance.

A coequalizer is the situation where there are distinct connections between the same two documents so that separate trails can pass across without merger. With the equalizer any separate trails arriving at the first document leave the second document by the same path. An equalizer is a context limit **C** represented in the diagram of Figure 8. All trails through **A** and **B** are merged through context **C** which will be shared by both **A** and **B**.

$$
\begin{tikzcd}
\mathbf{C} \arrow[r] & \mathbf{A} \arrow[r, shift left] \arrow[r, shift right] & \mathbf{B}
\end{tikzcd}
$$

Figure 8: The Equalizer **C** as a context limit on arrows from **A** to **B**

The corresponding coequalizer is given by Figure 9. The context of **C** is null, that is the limits are independent in thought but from the document perspective there is a context of documents where the two trails coexist with local independence. In other words links between documents may be equalizers or coequalizers.

$$
\begin{tikzcd}
\mathbf{A} \arrow[r, shift left] \arrow[r, shift right] & \mathbf{B} \arrow[r] & \mathbf{C}
\end{tikzcd}
$$

Figure 9: The Coequalizer **C** where two distinct trails coexist independently

Two other special limits are the terminal object and (its dual) the initial object. An object in a category **C** where there is one and only one arrow from every other object to it is known as the final or terminal object of $\mathbf{C}$. This may be denoted by $\top$ which is the last object in the trail. Dually (or oppositely) to the final object there may exist a corresponding initial object where there is an arrow from it to every other object in the category. This is $\perp$, the starting point in the trail and the arrows from it are every potential trail. This has significance for the reasoning and logical content that resides in the hypertext links.

In hypertext the initial and terminal objects may have only a local context. There may not be one single starting point, there may be a number of origins for any given trail. Likewise a trail may diverge to more than one finite point. Also natural language is a more general category than that of sets and the trails need not be disjoint. The same words could be used but with two distinct links in thought.

7 The Hypertext Lattice as a Heyting Algebra

As pointed out by Ted Nelson in his early idea of hypertext as "non–sequential writing with reader–controlled links" [Bolter 1990], links in hypertext are rarely linear but branch and form a distributive lattice. The internal logic of a lattice is geometric logic which is more general than Boolean logic. The logic of a lattice is well–established. It is equivalent to a Heyting algebra. Any Heyting algebra has a fundamental binary operation of implication $\Rightarrow$: $\mathbf{A} \longrightarrow \mathbf{B}$. This arrow is commonly written in the form $\mathbf{A} \Rightarrow \mathbf{B}$ and this shorthand version will be used here. This implication arrow is defined by the adjunction

$$\frac{(\mathbf{C} \times \mathbf{A}) \leq \mathbf{B}}{\mathbf{C} \leq (\mathbf{A} \Rightarrow \mathbf{B})}$$

$\mathbf{A} \Rightarrow \mathbf{B}$ is the largest category connected with $\mathbf{A}$ which is contained in $\mathbf{B}$. In hypertext terms if the current document ($\mathbf{A}$) in its context ($\mathbf{C}$) precedes document $\mathbf{B}$, then $\mathbf{B}$ is the next document after $\mathbf{A}$ in that context. In terms of concepts rather than documents, the concept may not be represented by a document in existence and from the point of view of a writer would be the next document to write.

By the application of this implication we can obtain the more generalized type of negation found in natural systems. Indeed in natural language it is often possible to represent negative concepts in a positive way. This is also true in hypertext where falsity and truth are not simple atomic entities. These are geometric concepts. Truth is given by $\mathbf{B} \Rightarrow \top$ and falsity by $\mathbf{A} \Rightarrow \perp$, sometimes written $\neg\mathbf{A}$.

Truth and falsity are relative to context. In hypertext, $\mathbf{A} \Rightarrow \perp$ is (usually back) in the direction of the initial document, a state of ignorance, whereas $\mathbf{B} \Rightarrow \top$ is forward in the direction towards the last document to be viewed in the lattice, the state of enlightenment. Knowledge and ignorance in hypertext are the counterparts of true and false.

The nature of the pseudocomplement then $\mathbf{A} \Rightarrow \perp$, that is not $\mathbf{A}$, may be further understood by substituting the special instance $\perp$ for $\mathbf{B}$ in the definition of the adjunction above. We then get

$$\frac{(\mathbf{C} \times \mathbf{A}) \leq \perp}{\mathbf{C} \leq (\mathbf{A} \Rightarrow \perp)}$$

In the real world two negatives do not always make a positive. This is familiar in natural language which opposes the principle of *tertium non datur*. The pseudocomplement is so important that natural languages often make it a separate word. For example the concept *relevant* has the pseudocomplement *irrelevant* which results in the further concept of *not irrelevant*. So *not irrelevant* is not equivalent to *relevant*. In fact there is a Heyting ordering:

	$\mathbf{B} \Rightarrow \perp$	$\leq (\mathbf{B} \Rightarrow \perp) \Rightarrow \perp$	$\leq \mathbf{B} \Rightarrow \top$
i.e.	irrelevant	$\leq$ not irrelevant	$\leq$ relevant

In hypertext terms, this gives a ranking of the relevancy of the documents in general terms for $\mathbf{B}$ the next possible document. It is an irrelevant document, if it is in the direction of the first document. It is the required next relevant document, if it is in the direction of the final document in the trail. Note when it is not irrelevant. That is, if it is not in the direction of the first document, whether or not it is in the direction of the final document. It is this three–level ordering which is the basis of much fuzzy thought and a generalization of fuzzy sets.

In terms of the Heyting algebra, $\mathbf{C} \Rightarrow \perp$ is another special case of $\mathbf{A} \Rightarrow \mathbf{B}$. As noted above $\mathbf{A} \Rightarrow \mathbf{B}$ is itself a concept/document and $\mathbf{C} \Rightarrow \perp$ is an *irrelevant* context concept/document. A fundamental feature is that the pseudocomplement $\mathbf{A} \Rightarrow \perp$ is the largest category disjoint from $\mathbf{A}$.

8 Geometric Consciousness

8.1 Contextual Awareness in Hypertext

The earlier discussion on context with pullbacks and pushouts deals with the simpler straight–forward type of static and objective contextuality but it is perhaps worth looking at the example previously raised:

"As I said in my speech on 28th October to this House ..."

A simple form of contextual awareness can be attained in this example by state of the art database techniques using fields, relations or keys. Thus the information identifying I, *House*, and *year* can be anaphorically resolved by reference to meta–records in the database system. Fuller details on how this works using keys in a Hansard database is given in [Rossiter & Heather 1988] and [Hudson 1985] where partial or composite keys are examples of colimits.

$$\begin{tikzcd}
year \arrow[r] \arrow[d] \arrow[dr, "\phi_2"] \arrow[drr, "\phi_1"] & \mathit{Hansard\ Record} \arrow[r] \arrow[dr, "\phi_3"] & Name \arrow[d] \\
28\ \ Oct \arrow[r] & House \arrow[r, "f"'] & I
\end{tikzcd}$$

Figure 10: Awareness to identify I and *House* in
"As I said in my speech on 28th October to this House ..."

The hypertext system of awareness is one that identifies for the user the next document to see. This is available from the implication $\mathbf{A} \Rightarrow \mathbf{B}$. Thus awareness is the contravariant natural transformation $\phi : \mathbf{B} \longrightarrow \mathbf{A}$. Awareness in hypertext is therefore the self identification of the document $\mathbf{B}$ in $\mathbf{A} \Rightarrow \mathbf{B}$. In Figure 6, it is the document records. The awareness to identify I is the *Hansard Record*, the *House* and the *Year* as given in Figure 10. This figure shows how the awareness works. The identity of the speaker I is given by ϕ_3, the identification of which house (Lords or Commons) by ϕ_2 and the awareness of the date of the speech from ϕ_1. These can be obtained algebraically. For example $f \circ \phi_2 = \phi_1$. f is the meta–record giving the house where the speech is given. In a database implementation f consists of those parts of the composite key which uniquely identify the *House*.

8.2 Computational Model of Consciousness

While the basic purpose of hypertext awareness is for that next document $\mathbf{B}$ in $\mathbf{A} \Rightarrow \mathbf{B}$ to identify itself to the user, the position is complicated by the fact that the user is operating at two levels the intension (represented by the document) and the extension (represented by the meaning). This identification is a precompositional contravariant arrow $\alpha^* : \mathbf{B} \longrightarrow \mathbf{A}$ which is a backward selection from the relevant documents of the document to which they are related.

Hypertext links are really connections between the semantic objects in the current document with related semantic objects in the documents to be retrieved. The connection is between objects $A_1, A_2, \ldots$ in the category $\mathbf{S}(\mathbf{A})$ (that is the meaning of the contents of the document under examination which is $\mathbf{A}$) with objects $B_1, B_2, \ldots$ in the category $\mathbf{S}(\mathbf{B})$ which are the meanings in the documents to be retrieved $\mathbf{B}$. The identification of the relevant documents depends upon the purpose and intentions of the user. This is a natural transformation $\eta_A : A \longrightarrow \mathbf{S}(\mathbf{B})$ as shown in Figure 11. The conscious awareness is given by the inverse natural transformation α^*.

$$\begin{tikzcd}
\mathbf{A} \arrow[r, shift left] \arrow[d, shift right, harpoon] & \mathbf{B} \arrow[l, shift left, "\alpha^*"] \arrow[d, shift right, harpoon] \\
\mathbf{S}(\mathbf{A}) \arrow[r, "\eta"'] \arrow[u, shift right, harpoon] & \mathbf{S}(\mathbf{B}) \arrow[u, shift right, harpoon]
\end{tikzcd}$$

Figure 11: Commuting Target Square for Awareness as a Natural Transformation

This is equivalent to analagous reasoning because the inverse natural transformation α^* preserves limits and colimits as well as implications.

$\alpha^*(B \Rightarrow B') \cong \alpha^*(B) \Rightarrow \alpha^*(B')$,
$\alpha^*(\top) \cong \top$,
$\alpha^*(B \times B') \cong \alpha^*(B) \times \alpha^*(B')$,
$\alpha^*(\bot) \cong \bot$,
$\alpha^*(B + B') \cong \alpha^*(B) + \alpha^*(B')$

The signs for products and sums are again used to represent generally limits and colimits respectively. Table 1 shows that none of the database relationships and structures in paragraph 4.2 require any operations beyond these. Therefore $\alpha^*(\mathbf{B})$ can claim to be a general awareness relationship.

An example of the reasoning counterpart to hypertext can be seen in the question discussed earlier: *Who pays a higher rent: tenants of controlled or uncontrolled tenancies?*. The inference that controlled tenancies pay lower rent (despite the apparent more valuable security of tenure) comes by Heyting logic. The geometric–logic representation is:

$$\frac{\mathbf{C} \leq (\mathbf{A} \Rightarrow \perp)}{(\mathbf{C} \times \mathbf{A}) \leq \perp}$$

If $\mathbf{A}$ represents a tenant paying rent, $\mathbf{A} \Rightarrow \perp$ represents the concept of a tenancy at low rent, from the simple two–object category $\mathbf{Rent}(\top, \perp)$. If $\mathbf{C}$ is a controlled tenancy, the proposition $\mathbf{C} \leq (\mathbf{A} \Rightarrow \perp)$ means that a controlled tenancy is equivalent to paying low rent. The effect of the control is that such rents are low. The inference $(\mathbf{C} \times \mathbf{A}) \leq \perp$ is that (the product) those who pay rent and are controlled tenants pay low rents.

8.3 Relative and Dynamic Contexts

Simple categories may be built–up to represent the greater complexity found in hypertext systems. For instance a concept that emerges from a structure of related documents itself is a diagram as previously indicated and may be used to replace a single object A. These can be employed to give hypertext the facility to deal with dynamic, subjective content. In geometric logic this amounts to manipulating more sophisticated structures for diagrams are a more general form of objects and simple categories.

For example the comma category has attracted considerable attention in computing science [Barr & Wells 1990] and can provide general contextuality. The comma category can add structure to an ordinary category by considering the arrows from the point of view of a particular object. Given a category $\mathbf{A}$ with a variable object A which may be represented by A' (when we want to distinguish different instances), the arrows $f : A \longrightarrow A'$ relative to C are objects in the comma category $\mathbf{A}/\mathbf{C}$ (sometimes written $\mathbf{A} \downarrow \mathbf{C}$) as shown in Figure 12. It should be emphasised that the objects in the comma category are arrows; the comma category arrows are triangles. For a map of the domain A and codomain A' together onto C specifies $f : A \longrightarrow A'$.

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr] & & A' \arrow[dl] \\ & C & \end{tikzcd}$$

Figure 12: Diagram of Comma Category

In practice hypertext does not just relate two documents but two documents in their respective categories. Therefore the hypertext link between document **A** (the one being viewed) and the next document **B** is given in Figure 13. The functor K is the hypertext link between the categories where the objects in each category are triangles composed of lower–level arrows. This shows up the dynamical aspects of context.

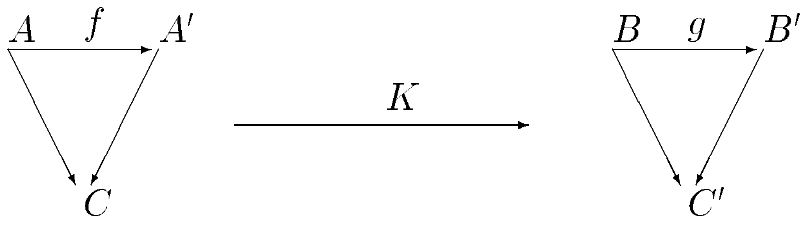

Figure 13: Covariant Functor K between Comma Categories

Figure 14 shows the corresponding contravariant functor α^* between comma categories.

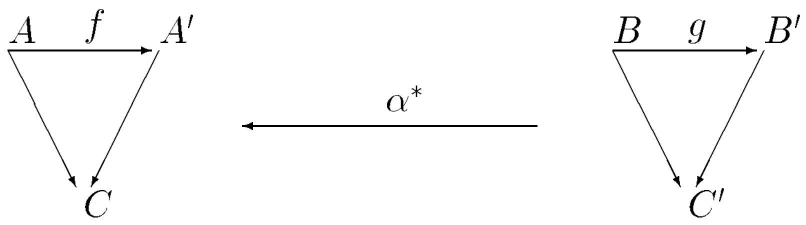

Figure 14: Contravariant Functor α^* between Comma Categories

Consciousness, with relative and dynamic context, is obtained by generalizing from the following relationships shown in Figures 13, 14:

$$K : \mathbf{A}/\mathbf{C} \longrightarrow \mathbf{B}/\mathbf{C}'$$

$$\alpha : f \longrightarrow g$$

$$\alpha^* : \mathbf{B}/\mathbf{C}' \longrightarrow \mathbf{A}/\mathbf{C}$$

The whole collection can be viewed as analagous reasoning thus confirming the equivalence of reasoning and hypertext.

9 Conclusions

The hypermedium has developed into a very sophisticated information system, one that is populated by a variety of distributed hypermedia source material. These heterogeneous materials are pulled together to create a contemporaneous document of the instant based upon the view of the user applied to the inherent structure in the source material. This view of the user is shaped by the real–world perceptions of the user interacting with the various documents encountered on the way. Because of the power to backtrack and because of inherent branches, it is more than just being shaped by a linear sequence of ideas.

If the system is to aid the user in handling this complexity, it is not only to show awareness of its own contents but must also be able to make the required inferences and connections. This is a form of consciousness. We have concentrated on a theoretical description in terms of geometric logic of the attributes of this machine consciousness as needed in information systems. This theory shows an equivalence of reasoning between documents and their contents. It seems therefore that this example from constructive mathematics may give us more insight into the concept of human consciousness. The general nature of the categories used in this study suggest that systems like the brain may operate in an analagous α^* natural transformation.

One certain result is that human consciousness is no less than machine consciousness and any theory of human consciousness must require at least the same capability to cope with concepts of geometric logic and to be able to operate at and across different levels. It is now recognized that a scientific representation of consciousness is needed for a *Theory of Everything* [Taylor 1993]. Constructive mathematics can perhaps provide an alternative [Barrow 1992] description of consciousness through geometric logic.

10 References

Barr, M, & Wells, C, *Category Theory for Computing Science*, Prentice–Hall (1990).

Barrow, J, *Pi in the Sky*, Oxford (1992).

Bing, J, & Harvold, T, *Legal Decisions and Information Systems*, Universitetsforlaget (1977).

Bolter, J D, Topographic Writing: Hypertext and the Electronic Writing Space, in: *Hypermedia and Literary Studies*, Delany, P, & Landow, G P, (edd.), MIT Press, Cambridge, Mass 105–118 (1990).

Chepalis, R, Documents at your Fingertips, *Law Society Gazette*, 7 December, **91**(45) 24-25 (1994).

Codd, E F, Extending the Database Relational Model to capture more meaning, *ACM*

TODS **4** 397-434 (1979).

Connolly, R, *The statute law and property database: an implementation of a full text data base on a relational data base management system*, M.Sc. Dissertation, Computing Laboratory, University of Newcastle upon Tyne, D313 (1985).

Freyd, P, *Abelian Categories: An Introduction to the Theory of Functors*, Harper and Row, New York (1964).

Freyd, P J & Scedrov, A, *Categories, Allegories*, North–Holland Mathematical Library **39** (1990).

Heather, M A, A Demand–driven Model for the Design of a "Half–intelligent" Common Law Information Retrieval System, in: *Computing Power and Legal Reasoning*, Walter, C. (ed.) 69–103 (1985).

Heather, M A, Cybernetic Consciousness, *Cybernetica* **XXXI**(1) 7–24 (1988).

Heather, M A, *Law as a Knowledge-Object Base*, Expert Systems in Law: an International Conference on Law & Artificial Intelligence organized by the University of Bologna on the occasion of its 9th Centennial. University of Bologna (1989).

Heather, M A, & Rossiter, B N, *Database techniques for text modelling: the document architecture of British statutes,* Computing Laboratory Technical Report Series no. 227, University of Newcastle upon Tyne (1987). (66pp)

Heather, M A, & Rossiter, B N, An Interoperable Theory of Higher Level Logic and Integrable Law for the Environment: In Celebration of the 25th Anniversary of the Founding of the IDG in: *Towards a Global Expert System in Law*, Florence 1993, Cedam, Padua (1993).

Heather, M A, & Rossiter, B N, Applying Geometric Logic to Law, in: *4th National Conference on Law, Computers and Artificial Intelligence*, Exeter 1994, 80–95 (1994).

Hudson, G, *Establishment of a data base containing the information of Hansard using a hierarchical data base management system SPIRES*, Masters Dissertation, Computing Laboratory, University of Newcastle upon Tyne, D289 (1985).

Lambek, J, & Scott, P J, *Introduction to Higher Order Categorical Logic*, Cambridge (1986).

Lawvere, F W, Adjointness in foundations, *Dialectica* **23** 281–296 (1969).

Mac Lane, S, *Categories for the Working Mathematician*, Springer–Verlag, New York (1971).

Manager, J J, *The World–Wide Web*, Mosaic and More, Mcgraw–Hill (1995).

Marovac, N & Osburn, L, HyperNet – A Tool to Choreograph World Wide Distributed Hypermedia Documents, *Computers & Graphics*, Pergamon Press, Oxford (1992).

Nelson, D A, & Rossiter, B N, Prototyping a Categorical Database in P/FDM, in: *ACM Workshop ADBIS 95*, Moscow, June 1995, 247–258 (1995).

Nielsen, J, *HyperText & HyperMedia*, Academic Press, New York (1990).

Rhydeheard, D E, & Burstall, R M, *Computational Category Theory*, Prentice–Hall (1988).

Rossiter, B N, Machine awareness in database technology, Proceedings Symposium VI, Meta-intelligence and the Cybernetics of Consciousness. *XI International Congress of Cybernetics,* Namur, 1-9 (1987).

Rossiter, B N, & Heather, M A, Data models and legal text, *CC AI* **5**(1) 39-55 (1988).

Rossiter, B N, & Heather, M A, *Towards the Object-oriented Textbase,* Computing Laboratory Technical Report no.297, University of Newcastle upon Tyne (1989). (49pp)

Rossiter, B.N. & Heather, M.A, Strengths and Weaknesses of Database Models for Textual Documents, *Proceedings EP90*, ed. Furuta, R. Cambridge 125–138 (1990).

Rossiter, B N, Nelson, D A, & Heather, M A, *The Categorical Product Data Model as a Formalism for Object–Relational Databases*, Computing Science Technical Report no.505, University of Newcastle upon Tyne (1994) (42pp).

Rossiter, B N, Sillitoe, T J, & Heather, M A, Database Support for Very Large Hypertexts: Data Organization, Navigation and Trails, *Electronic Publishing - ODD,* **3**(3) 141-154 (1990).

Rymer, J, Guttman, M, & Matthews, J, Microsoft OLE 2.0 and the road to Cairo, how object linking and embedding will lead to distributed object computing, *Distributed Computing Monitor*, Jan (1994).

Spaccapietra, S, ed. *Entity-Relationship Approach: Ten years of experience*, North Holland (1987).

Taylor, J, *When the Clock Struck Zero*, Picador 1993.