Dagobert Soergel January 16, 2003

Representing knowledge about databases in an intelligent guide to databases

Contents

1. Introduction

With the large number of databases available today, even experienced searchers cannot rely on their memory for database selection; for them, too, this step has become a retrieval problem of its own. In terms of just the numbers of items involved, this retrieval problem may seem much less formidable then retrieval in a large bibliographic database such as Chemical Abstracts. However, many factors must be considered when judging the usefulness of a database for a search. Furthermore, for online databases one must take in to account the characteristics of the search system on which they are implemented and the users familiarity with this search system.

The factors to be considered are so many and the connections and computations to be made so complex, that the human searcher cannot cope and cannot be expected to make optimal selections with any consistency. Furthermore, many requests require the use of several databases in combination, adding another layer of complexity to the selection process. An expert system, or knowledge-based retrieval system, is needed for a satisfactory solution of these problems.

2. The functions of a database guide in a search

A database guide should be seen in the context of a comprehewnsive systems concept as illustrated in Figure 1. The user enters a free-form statement of the problem and rteh resulting information need; this stastement may be minimal (e.g., just one word) or, in the extereme case, empty. This free-form statement is interpreted, augmented, and transformed into a query formulation by the query analysis component. This component may be a humna intermediary, a program, or a combination of both.

This paper deals with one aspect of such a system, namely representing knowledge about databases. It starts with a sketch of the architecture of an intelligent database guide and then proceeds to a detailed discussion of the elements needed in the representation of knowledge about databases. Put differently, it develops a conceptual schema for database description.

While this paper focuses on the description of reference tools in print or on-line form, it has implications for description of any kind of database.

The ideas set forth in this paper were developed in a small pilot project conducted at the University of Konstanz with a small grant from the German Ministry for Research and Technology. A small database was produced and a prototype search program was written. But this work served mainly to develop ideas on which a systems development effort can be based.

3. Functions and structure of an intelligent guide to databases

3.1 Functions

The intelligent guide to databases to be described in this paper serves the following purposes:

- Retrieval of databases that should be searched - alone or in combination - for a given request; this occurs at the beginning of a search and during a search if the search in the first database suggested is not successful.

- Assistance to the user doing a search in one database and in the transformation of a query formulation from one database to another.

- Production of printed documentation about databases. This includes documentation on one database in general (such as Chemical Abstracts), comparative documentation (comparing several databases in a subject area), documentation on the implementation of a database on a given search service (Chemical Abstracts on DIALOG), and comparative documentation comparing several implementations of a database.

3.2 Design Alternatives

There are two ways to accomplish intelligent retrieval. Alternative 1 is a straightforward rule-based system with rules that connect query characteristics directly to databases. One kind of rule would have form

IF combination of search request characteristics

THEN recommend database X with strength Y,

where query characteristics include subject, type of data needed, level of use, and urgency.

For example

IF query topic Legislation
AND selection by Subject
AND information needed Bill status in Congress
THEN recommend LEGISLATE with strength .9

The system has no knowledge as to why this database should be selected.

Another kind of rule would be:

IF database X recommended
AND Database X implemented on search system Y
AND user knows search system Y

THEN recommend search system Y

This approach requires a large number of rules; it would be impractical to have enough rules to do justice to the complexity of the selection problem. Consider the variety of search requests and the interaction between the content of a database and the way it is implemented on a particular search system. Furthermore, with every new database many new rules must be added.

In alternative 2 the system is empowered to draw conclusions from descriptive data on databases and search requests. In a Prolog implementation, descriptions of databases and search requests are stored as a collection of Prolog facts. Prolog rules provide for reasoning from these facts. This approach is much more general and powerful. Adding a new database requires adding all pertinent facts - not a small task by any means but a task requiring considerably less effort than adding a large number of rules.

To achieve good retrieval results, the database descriptions should be rather detailed. But these detailed descriptions also serve for user assistance during the search and preparation of documentation, functions that are not supported in the strictly rule-based approach.

To sum up, in alternative 1 rules are specific, connecting specific search situations with individual databases. Knowledge about databases is encoded in rules. In alternative 2, rules are general. The rules encode knowledge of how to deduce requirements for databases from search request description. Knowledge about databases is encoded as a collection of facts.

The system architecture proposed here is based on the facts-and-reasoning approach. However, some specific rules of the type used in alternative 1 are useful. They can be recorded as facts about databases and used in the reasoning about database selection like any other fact.

3.3 System architecture

It is clear from the previous section that description of databases and search services and of user profiles and search requests play a central role in the design of a database guide. Therefore, conceptual schemata for description require considerable thought.

This section briefly describes the architecture of the database guide (see Figure 1) setting stage for the detailed discussion of the conceptual schema for database description to be given in the remainder of the paper. We first sketch the structure of the knowledge base and then the components of the program to build and search the knowledge base.

3.3.1 The structure of the knowledge base

*** revise to take account of new 4. and 5. plus old 4/5 is now 6/7

The knowledge base consists of data about databases and search services and data about users and search requests. Data about databases cover such topics as availability and cost, audience and coverage, database structure (database content and retrieval access), and user support provided.

The content of a database is defined by the recording of its conceptual schema in accordance with the entity-relationship approach. Entity types and relationship types are standardized across databases. (A structured list of entity types and relationship types is being developed, with databases analyzed serving as sources; the raw list is given in the appendix.) Thus the data about databases can be divided into a general part - data about entity types and relationship types - and a special part - data about individual databases and their implementations. The general part contains also thesauri; each thesaurus contains entity values for a given entity type, or possibly for several entity types. In particular, there is a thesaurus for the entity type concept/subject. Such a thesaurus must contain at least the concepts needed to define database coverage. Preferable is a very large, hierarchically structured thesaurus that would allow mapping from - usually specific - query concepts to - usually more general - concepts defining database coverage. As it turns out, it is useful to include the definition of selection criteria in the general part, whereas the application of these criteria to describe the coverage of the individual databases belongs to the special part.

Thus the knowledge base of the database guide consists of five components:

Data about databases and search services

1. General data about entity types and relationship types; thesauri; definition of selection criteria.

2. Data about databases and their implementation.

3. Data about search services.

Data about problem-solving and search strategies

4. Data about problems, problem-solving strategies, and connections between problems and information needs.

5. Data about search strategies in general.

Data about users and search requests.

6. User profiles, including user interests and subject knowledge and user knowledge of and access to databases, search services, and search programs.

7. Data about search requests.

3.3.2 Structure of the program for building and searching the knowledge base

*** revise as above

For each of these knowledgebase components there must be a program for the acquisition of knowledge/data. The user interacts with the acquisition program for user profiles and search requests. The acquisitions program for search requests produces a formal representation of all aspects of the search request.

The search program uses all these data in the retrieval of databases suitable for a search request. It consists of 3 modules. All modules use the data about user profiles and the data about search requests. But with respect to data about databases and search services, each module specializes on one knowledge component as follows.

MODULE 1 matches the query formulation with knowledgebase component 1 (general data about entity types and relationship types) and produces a list of requirements for suitable databases in terms of the entity types and the relationship types covered in their conceptual schema and in terms of selection criteria.

MODULE 2 matches this list of requirements with knowledgebase component 2 (specific data about databases) and produces a list of suitable databases and their implementations. For each database an estimate of its usefulness in answering the search request is derived. This module should also take into account data about search strategies in general.

MODULE 3 uses knowledgebase component 3 (data about search services) to evaluate each

database implementation in terms of accessibility for the user and in terms of costs and derives a cost-benefit ratio. Only databases above a given cut-off point will be displayed.

The display program shows the user a well-organized list of the databases found with appropriate information, such as usefulness, cost, search service, database number (number on the search service). This program would also take care of formatting data for printed documentation.

There are many interactions between these program parts. For example, the system may do a rather superficial analysis of the search request and a superficial pass through the search program, and come up with a database that is easily accessible to the user. If the search in this database is not successful, the process starts all over again and is executed more thoroughly. To give another example, it is usually advisable to ask the user certain questions only if the answer to these questions make a difference in the selection of databases; often this can be determined only during execution of the search program, which then must make a request to the search request acquisitions program, which in turn will ask the user. The database guide system needs an executive program that controls these interactions between the various program parts.

4. The conceptual schema for database description

The description of databases is at the heart of a database guide; the descriptions must represent the information needed for searching in a concise manner. There is much to know about databases and the development of a conceptual schema entails many decisions about how to represent this knowledge. This section gets, by necessity, into the nitty-gritty detail of database description. The conceptual schema was developed from an analysis of the literature on database selection and on database description, and it was refined while describing 15 databases and some of their implementations through specific search services, such as DIALOG and Data Star.

One may object to the conceptual schema presented here on the grounds that the data it allows to represent are not available or would be too costly to compile and input. However, while the schema allows for detailed description, it does not mandate it. The amount of information included, and therefore the quality of the database selection and user assistance, is a matter to be decided for each implementation. Much information is available in documentation provided by database producers and search system vendors. But a great deal of information, for example about coverage , overlap, quality, and search experience, is scattered in the literature, and a database guide could provide - at considerable cost - an invaluable service by bringing all this information together in a codified form.

In our conception of a database guide, each database description is a set of Prolog facts. Each Prolog fact is a list of entity types which are in a named relationship. Prolog is a natural for implementing the entity-relationship approach.

The conceptual schema to be described is, then, a list of entity types and relationship types. The text introduces these entity types and relationship types in an order that is most suitable for understanding. Many relationship types are illustrated by examples from the description of MEDLINE. The MEDLINE description (as far as it is developed) and the full conceptual schema in a sequence that by and large goes from the general to the specific are given in the appendix. The reader may find it helpful to consult their description while reading the discussion of the conceptual schema. For ease of understanding, relationships are given in a form that is close to natural language; the relationship name appears between two arguments, where each argument can be an entity or a list of entities. In many cases an example immediately follows the general form of a relationship.

4.1 Relationships concerning availability and costs

4.1.1 Formal data

We start with some straightforward relationships; db stands for database.

db **has-name** name

MEDLINE has-name MEDLARS

Each database is identified by a standard abbreviation. This relationship gives names other than the standard abbreviation.

db **has-short-description** text

db **is-of-type** db-type

db-type has sample values rule-work, thesaurus, LIST (value-1...value-n) Other values are not assigned at present

db **produced-by** legal-ent

MEDLINE produced-by NLM

db **sponsored-by** legal-ent

This relationship is used when a database is produced by one person or organization on the direction, funding, or authority of another person or organization.

4.1.2 Subdivisions of a database: subset, member, implementation, segment

The following relationship types require a more differentiated view of the entity type **database** There are no values associated with the general entity type db itself; the entity values belong to one of the following subtypes:

Database - general	db-gen	Refers to database content at the conceptual level. Example: MEDLINE
Database implementation	db-impl	A particular implementation of the database (online, print, etc.), which records the content of the database and makes it accessible and searchable. Example: MEDLINE-DL, i.e. MEDLINE as searchable on DIALOG.

If db is indicated in the format of a relationship type, the value entered can belong to any subtype.

With these definitions we can now introduce relationships between databases.

db-1 **contains** db-2

An unspecific relationship, covering **contains-subset**, **contains-segment**, and **contains-member**

db-1 **contains-subset** db-2

MEDLINE contains-subset IDL (Index of Dental Literature)

A subset of a database, defined by some criterion other than date of entry or publication date of a db-impl. A subset may be extracted from a database for marketing to a special audience. Or a subset may be a part of a database that is generated by a separate producer and integrated into the larger database.

db-1 **contains-segment** (db-2 date-1 date-2)

MEDLINE-DL contains-segment (DL-153 1973 1979)

A segment is a subset of a database defined strictly by entry date; for example, DL-153 (DIALOG file 153) contains records that have been entered in MEDLINE in the years 1973-1979. Contrast with the subset relation.

If db-1 is a db-gen, so is db-2; if db-1 is a db-impl, so is db-2

Beginning and ending date can be defined relative to TODAY:

db-impl **contains-segment** (db-segm (TODAY - m3) TODAY)

defines a segment containing records from the last three months. With monthly update, this means that, as a new month is added, the oldest month is deleted.

db-1 **contains-member** db-2

This relationship is used to record the structure of a relational (or semi-relational) database, where the members are the individual relations (tables). Each relation (table) is considered of entity type db (db-gen or db-impl, as the case may be). If one interprets this broadly enough, it includes the structure of a computerized or printed implementation, the various files (including index files) being the members. The thesaurus used in a database is always a member, but this relationship should be recorded only for on-line databases for which the thesaurus can be used in searching in a relational mode, as in the EXPLODE command in MEDLINE.

contains-member should not be confused with **contains-subset**. A subset is a database of the same structure as the parent but with more limited coverage. A member of a database, on the other hand, contains information that is or can be used in conjunction with other members.

see also the relationship

db-1 **has-index** (((rel-1...rel-n) db-2) db-3)

db-gen **implemented-as** (db-impl doc-form legal-ent title)

MEDLINE implemented-as (MEDLINE-DL online-chardata DIALOG $)

The entity type **doc-form** (document form) has values such as online-chardata, online-facsimile, cd-rom-chardata, and hc (hard copy). **legal-ent** is the person or organization producing or marketing the db-implementation. **title** is used only when otherwise the data about various segments would become very redundant. Here title is not needed; therefore $.

Any subset, segment, member, or implementation can in turn be described using the full conceptual schema. Much of the information is implied by hierarchical inheritance: many statements that are true of a database are also true of its subsets, segments, and implementations. Only statements that are not implied by hierarchical inheritance need be recorded explicitly. The precise rules for hierarchical inheritance must be recorded in the database guide. Hierarchical inheritance makes for more efficient data entry and storage.

4.1.3 Accessibility and costs

The following relationship types are used to record data about accessibility and costs. Since these are often dependent on the class of users, a new entity type, class-legal-ent, is introduced. Sample values are physicians or subscribers-to-db.

db **accessible-for** class-legal-ent

(db-impl class-legal-ent) **db-specific-connect-hour-cost** ((money-number currency)...)

(db-impl class-legal-ent) **on-line-display-cost** ((money-number currency)...)

(db-impl class-legal-ent) **downloading-cost** ((money-number currency)...)

Downloading is transfer to a permanent machine-readable file. Temporary machine-storage for making a printout is not downloading. When downloading is not permitted, use a money-number of all "9".

(db-impl class-legal-ent) **off-line-print-cost** ((money-number currency)...)

4.2 Relationships to describe audience, number of entities, updating, sources used, and miscellaneous characteristics

4.2.1 Audience and purpose

db **intended-for** (class-legal-ent purpose)

MEDLINE intended-for (medical-researchers $)
MEDLINE intended-for (physicians patient-care)

db	**suitable-for**	(class-legal-ent purpose rating)
MEDLINE	suitable-for	(medical-researchers $ 1)
MEDLINE	suitable-for	(physicians patient-care 2)
MEDLINE	suitable-for	(physicians continuing-ed 2)

Ratings go from 1 (A) (best) to 6 (F) (worst). No statement is entered if the rating is 4 or worse. If either audience or purpose are not specified, use $; this makes it possible to record suitability for an audience, a purpose, or a combination.

db	**used-by**	(class-legal-ent purpose rating)

suitable-for serves to enter judgments by the author of a statement. This relationship serves to record data uncovered in user studies; the rating would be an average derived from users' answers.

4.2.2 Number of entries and updating

*** This has problems in relation to coverage. Also contains-segment and update.

db	**has-total-entries**	(entity-type date number)
MEDLINE	has-total-entries	(document 1982.04 3750000)
MEDLINE	has-total-entries	(chemical-substance ? ?)
DL-154	has-total-entries	(document ? ?)

entity-type is included explicitly since one may wish to record the number of values for an entity type other than the "head" entity type; for example, it is not an unreasonable question to ask how many substances are dealt with in documents covered by MEDLINE.

The **updating pattern** of a database, its implementations and segments, can be rather complex; there may be cumulations, supplements, supplements to supplements. We define two relationships which, in conjunction with the relationship **contains-segment**, are sufficient to describe most updating situations.

These relationships should be given for a database in general (db-gen) and for all database implementations, normally the affected segment (db-segm), which may be a supplement containing only the current month, or the current half-year, etc.

db	**is-updated-time**	(date time-span delay)
MEDLINE	is-updated-time	(1982.04 m m6)
SDILINE	is-updated-time	(1982.04 m m6)
DL-154	is-updated-time	(1982.04 m w2)

Date is the first date, from which all subsequent dates can be computed using time-span.

time-span is given by a code number such as m (monthly), m/2 (every half month = twice a month), m2 (every two months). For a db-gen, date and time-span refer to the issuance of updates by the database producer; for a db-segm, they refer to the incorporation of the changes.

Delay. For a db-gen, delay, in days, is the time elapsed between the appearance of an entity and its inclusion in a db-gen. For a db-segm, delay is the time elapsed between availability of the last batch of data from the database producer and completion of the update. For example, if the database producer sends data weekly and update is done monthly, delay is the time between availability of the last weekly batch and completion of the update.

db	**is-updated-numbers**	(entity-type number-of-additions number-of-deletions number-of-corrections)
MEDLINE	is-updated-numbers	(journal-article 20000 0 0)
SDILINE	is-updated-numbers	(journal-article 20000 ALL 0)
DL-154	is-updated-numbers	(journal-article 20000 0 0)

number-of-additions means number of added records (or tuples of a relation).
number-of-deletions can take the value ALL: all old records in a database or a database segment are deleted. This applies, for example, to a database implementation that maintains a segment of just the last update for SDI searches.

4.2.3 Sources

db-1	**uses-source-direct**	(db-2 extent)
MEDLINE	uses-source-direct	(ISSN-235784 c)
db-1	**uses-source-indirect**	db-2
MEDLINE	uses-source-indirect	MEDLINE-journal-list

A **direct source** is one from which entries for the database are taken directly, for example, a bibliography or journal from which entries are taken for a bibliographic database. A journal is identified through its International Standard Serial Number (ISSN). The entity type extent has, as of now, two values: completely (all articles from the journal are included in MEDLINE) and partially (only some articles are included). A more detailed code would record complete or partial inclusion by type of item, such as full article, brief report, book review, etc.)

An **indirect source** is one that in turn contains sources for the database (for example, a list of journals, each of which serves as a source from which articles are taken for a bibliographic database).

Sources are themselves databases that can be described using the total conceptual schema; at least some facts about them, especially concerning coverage, should be recorded. The use of a source often implies a selection criterion, and this must be considered in searching.

Relationship types for recording selection criteria, coverage, and overlap would logically be described here. However, they are rather complex, and we defer their treatment until later when all necessary prerequisites are available.

4.2.4 Miscellaneous

legal-ent **has-qualification** (criterion rating)

legal-ent **has-bias** (bias rating)

Consider, for example, a database on Health effects for smoking produced or sponsored by the American Tobacco Institute.

db **has-special-service** special-service

Sample values of special-service are document-delivery-service and comment-service. (For example, in the IRCS database of full-text medical papers users can enter comments while are then incorporated in the record)

db **has history** text

The following relationships are used mainly for databases that are thesauri or other vocabularies or authority lists or rule works.

db overall-quality ((subject-1 rating-1) (subject-2 rating-2) ...)

db descriptor-selection ((subject-1 rating-1) (subject-2 rating-2) ...)

db specificity ((subject-1 rating-1) (subject-2 rating-2) ...)

db conceptual-structure ((subject-1 rating-1) (subject-2 rating-2) ...)

db facet-analysis ((subject-1 rating-1) (subject-2 rating-2) ...)

db hierarchy ((subject-1 rating-1) (subject-2 rating-2) ...)

db hierarchy-display ((subject-1 rating-1) (subject-2 rating-2) ...)

db RT-crossreferences ((subject-1 rating-1) (subject-2 rating-2) ...)

db descriptor-names ((subject-1 rating-1) (subject-2 rating-2) ...)

db lead-in-vocabulary ((subject-1 rating-1) (subject-2 rating-2) ...)

These relationships allow for the recording of various aspects of thesaurus quality, with several ratings for an aspect qualified by subject, if necessary; for example

MeSH specificity (($ 1) (Medicine 1) (Economics 4))

The first rating is a general rating across all subjects.

db exercises-control control-level

Sample values of control level are: authority-control, completely-enumerated, combinations-from-given-elements, free-with-form-constraints, free

db **has-main-language** language

For a thesaurus, language is the language of the descriptors. If this thesaurus is used in a database, the thesaurus language is the language of the descriptors used in the database. To give another example, the language to be entered for a rule-work for abstracts is the language in which the abstracts are to be written.

4.3 Relationships for recording the conceptual structure of a database and retrieval access

4.3.1 Introduction

The conceptual structure of a database is described by giving its conceptual schema in accordance

with the entity-relationship approach. This works for any type of database, even if the database is structured into records, as is the case for most bibliographic databases. A record can be seen as a set of statements about a "head entity" (In a bibliographic database, **document** is the head entity type; in a biographical database, **person**; in a job listings database, **job**). Each data field in the record corresponds to a relationship type. A list of relationship types is thus equivalent to a list of data fields. To describe a relational database, each individual table (relation) is treated as a database for purposes of description. The fact that a database consists of several tables is expressed through the **contains-member** relationship discussed above.
The entity types and relationship types used in recording the conceptual schema of a database are standardized across databases. In other words, there is a thesaurus of relationship types. Some sample entries are:

rel-002 is-rel (entity is-about-or-relevant-for subject)

rel-008 is-rel (document authored-by person)

rel-013 is-rel (document emanated-from organization)

rel-032 is-rel (person is-affiliate-with organization)

rel-033 is-rel (legal-ent located-at address)

rel-071 is-rel (legal-ent located-at city)

The entries in the thesaurus of relationship types refer to a relationship connecting two entities of a given type. The left hand entity type is often very general (in the most general case simply "entity"). In the description of a specific database it is understood that the left hand entity type of any relationship is restricted to the head entity type of the database. On the other hand, the right hand entity type is usually quite specific. For example, rel-008 specifies person and rel-013 specifies organization. In this way it is possible to distinguish the data field that gives a personal author from a data field that gives a corporate author.

So far we were able to restrict ourselves to binary relationships. Generalization to n-ary relationships would not pose a major difficulty; the format of the **gives-data-on** relationship to be discussed below would need to be generalized somewhat.

4.3.2 Domain and other statements about entity types

db-1 **controls** (entity-type identifier-form db-2)

db-1 is the database using the thesaurus or rule work.

Sample values of **identifier-form** are name, term, phrase, text, string, acronym, code, number, range-code, address.

db-2 is a thesaurus or rule work that controls the entity values; it is a database of its own, and can be described using the full conceptual schema. For example, a minimal description of MeSH (Medical Subject Headings, the thesaurus used in MEDLINE) consists of two statements:

MeSH is-of-type thesaurus
MeSH exercise-control combination-of-given-elements

MeSH has-main-language eng
(indicating that MEDLINE uses a controlled vocabulary, the language of which is English.)

The type of db-2 shows whether control is by rule or by authority list. If thesaurus or rule work not known, use T or R, respectively. If no control at all, use F-all for db-2. T, R, and F-all are pseudo databases with very brief descriptions:

T is-of-type thesaurus
T exercises-control authority-control
T has-main -language nsp (not specified)

R is-of-type rule-work

F-all is-of-type rule-work
F-all exercise-control free
F-all has-main-language all

After db-2 any of the following pairs may be given to represent more detail:

(CSR consistency-rating))

In the argument list, these elements are recognized by key word rather than position.

consistency-rating indicates how consistently control is enforced

controls is used rarely; usually this information is given as part of an **gives-data-on** statement (see below). **controls** is used only to avoid redundancy when the same entity type with the same control is used in many relationships.

db-1 **has-search-help** (entity-type identifier-form db-2)

A search help is a thesaurus (in the most general sense) that is not used for control in indexing but is useful to a searcher in locating terms and other identifiers to be included in the search.

db **treats-as-subtype** (entity-type-1 entity-type-2)

MEDLINE treats-as-subtype (disease subject)

This relationship indicates that a search for databases dealing with **disease** should find MEDLINE even though **disease** is not explicitly mentioned in a gives-data-on statement for MEDLINE. MEDLINE also treats the entity types **organism**, **body part**, **chemical substance**, and **enzyme** like **subject**.

db **treats-as-component** (entity-type-1 entity-type-2 percent/rating)

IRCS treats-as-component (phone-number address 70)

4.3.3 Data about relationships appearing in a database: database level

db	**has-head-entity-type**	entity-type
MEDLINE	has-head-entity-type	journal-article
Who's Who	has-head-entity-type	person

Before we formally introduce the relationship **gives-data-on**, we give a few examples that illustrate its use. To save space, the relationship types are referred to by their identifiers from the relationship thesaurus, such as rel-002, rel-008, etc. (see above).

MEDLINE gives-data-on ((rel-002) term MeSH)

The database MEDLINE gives data on the relationship type

entity is-about-or-relevant-for subject;

the subject is identified by a term, and the possible values of these terms are controlled by MeSH (Medical Subject Headings). (The right hand side of **gives-data-on** has the same format as the right-hand side of **controls**.)

Another example is

MEDLINE gives-data-on ((rel-008) name R)

MEDLINE gives data on the relationship

document authored-by person

document, the left hand element of rel-008, is restricted to the head entity type of MEDLINE, journal article. **person**, the right hand element of rel-008, is identified by a name, which is governed by unspecified rules. (See the explanation for controls.)

A MEDLINE record gives also data about the person, specifically the organization where the person works (the database is clearly not normalized in this instance). This situation occurs quite often, and we have developed an elegant way to deal with it using a chain of relationships:

Document authored-by person / person affiliated-with organization

Thus the **gives-data-on** statement becomes

MEDLINE gives-data-on ((rel-008 rel-032) name R)

This relationship chain makes it very explicit that MEDLINE can be used for finding information about the organizational affiliation of a person. When asked for databases containing information about organizational affiliation, the search program first looks for **gives-data-on** statements with rel-032 as the first (and often only) element of the relationship chain. If this does not lead to a suitable database or if the search in a database recommended does not succeed, the search program then looks for **gives-data-on** statements that contain rel-032 anywhere in the relationship chain and thus finds MEDLINE. In this search the system should also consider that the affiliation

information in MEDLINE may be more up-to-date than in a biographical database.

In order for the chain to be properly constructed, the element on the right hand side of rel-008 must be the same entity type as the element on the left hand side of rel-032 or a subtype of it. The element on the right hand side of rel-032 is recorded in the MEDLINE data field using a name governed by unspecified rules.

Chains can be longer, as in the following example

MEDLINE gives-data-on ((rel-008 rel-032 rel-071) name R).

This means that MEDLINE also gives the city of the author's organization. Thus, MEDLINE can also be used in a search for the city where an organization is located. It can be easily used this way if the implementation provides an index by organization. But even if no such index is given, not all is lost. If we know a person affiliated with the organization, we can search for documents authored by that person; any of the document records found will show the organization and its city.

Another example of a chain is

MEDLINE gives-data-on ((rel-002 rel-095) code MeSH)

It says that the subject expressed as a MeSH **term** may also be expressed as a MeSH **code**. It is the same concept in both cases, but the code is found in a different data field as indicated in a **records** statement.

Use of a chain always indicates that the database being described is not normalized with respect to the relationships involved.

We can now state the general form of the **gives-data-on** relationship

db-1 **gives-data-on** ((rel-1 ... rel-n) identifier-form db-2)

db-1 is the database in which the relationship type appears. This is normally a db-gen, such as MEDLINE, rather than a db-implementation, such as MEDLINE-DL. Occasionally a search system like DIALOG augments a database, and the **gives-data-on** relationship then refers to the specific database implementation as in the following examples:

MEDLINE-DL gives-data-on ((rel-036) seqno R)

where

rel-036 is-rel (entity has-db-impl-accession-no accession- number)

or

MEDLINE-DL gives-data-on ((rel-023) code DIALOG-doc-1)

where

rel-023 is-rel (entity is-about-or-relevant-for subject-pre- ex
pl

od
ed
)

This means that DIALOG has pre-exploded some MEDLINE subjects; these subjects are represented by a code that can be found in a document known as DIALOG-doc-1.

The first argument on the right hand side of **gives-data-on** is a **relationship chain**.

Identifier-form is the form of the identifier used for the right hand entity type of the last relationship in the chain. Sample values for identifier-form are name, term, text, "mm.yy" (for a months given in the form 04.86), code, address.

db-2 is the database (or pseudo-database) governing the admissible values of the identifier.

There may be two **gives-data-on** statements with the same relationship chain for one and the same database. For example, the subject for which an entity is relevant (rel-002) may be given according to two different thesauri, as in Excerpta Medica, which uses two index languages, MALIMET and EMCLASS; the distinction is maintained through db-2:

EM gives-data-on ((rel-002) term MALIMET)
EM gives-data-on ((rel-002) code EMCLASS)

After db-2 any of the following pairs or triplets may be given to represent more detail:

(CS consistency-percent)
(CSR consistency-rating)
(CONTENT-RULES db-3)
(SEARCH-HELP db-4)
(C completeness-percent)
(CR completeness-rating)
(E entity-value exhaustivity)
(ER entity-value exhaustivity-rating)
(S entity-value specificity)
(SR entity-value specificity-rating)
(V viewpoint)

Note. In most relationship formats defined here the meaning of a piece of data is determined by its position in a list. This works fine as long as all values in the list are present most of the time and none occurs twice. Neither one of these conditions is fulfilled here. Therefore, each value is represented as a list; the first element of the list indicates the meaning of the value.

consistency-rating see explanation for controls

db-3 is a work of rules that specify when an indexer should introduce a relationship of the type given.

db-4 is a thesaurus or other list of entity values and lead-in designations that is not used to control entity values entered in the database but is available as a search help. If this thesaurus can be used for automatic query expansion, it should be recorded as member database of db-1.

Completeness refers to the number of statements of the type defined by the relationship chain that

are included in the database in relation to the number that should be.

Exhaustivity and **specificity** are used in the generally accepted definition. The entity-value slot can also be filled with a description of a class of entity values. This allows for differentiated statements, such as high specificity for science concepts but low specificity for economic concepts. Statements on exhaustivity can in principle be made for each individual entity value (e.g. each individual subject descriptor) in the database; such information would indeed be most helpful in determining which database is best for searching a particular subject.

Viewpoint applies primarily to abstracting and subject indexing. For example, one may analyze medical documents with the interests of a pharmaceutical company in mind, or documents in theoretical physics from the point of view of a philosopher. Several viewpoints may be used simultaneously.

4.3.4 Data about field names and retrieval access: database implementation level

The following relationship, **records**, is used to indicate that a relationship is recorded in a database implementation and to specify, in a very compact form, information about field names and retrieval access. (Not all relationships for which a database gives data are recorded in every implementation; for example, Chemical Abstracts as implemented on DIALOG or any other system except CA's own search service does not record the abstracts of documents.) If the same statement is true for all or most implementations of a database, it should be given at the db-gen level to save space.

There must be a **gives-data-on** statement for every **records** statement. In rare cases a database implementation provides more than one mode of retrieval access for a data element corresponding to a **gives-data-on** statement; in such a case, there is one **records** statements for each mode of access.

db **records** ((rel-1 ... rel-n db-2) field-name-1
retr-access/sort-sequence field-name-2 field-name-3
LIMIT LIMIT-SELECT (VALUES value-1 ... value-n) RANGE
NO-DISPLAY SORT)

We first give some examples and then explain the individual components of this relationship.

Examples:

MEDLINE-D records ((rel-002) DE wr "")
MEDLINE-DL records ((rel-002 rel-095) DC rr)
MEDLINE-DL records ((rel-008) AU rr SORT)
MEDLINE-DL records ((rel-008 rel-032) CS rr)
MEDLINE-DL records ((rel-024) $ r LIMIT-SELECT (VALUES human nonh))

The **relationship chain** is the same as that used in the corresponding gives-data-on statement. db-2 is given only if there are two **gives-data-on** statements with the same relationship chain, as in the case of Excerpta Medica which gives the subject according to two different thesauri; each subject indication has its own field.

field-name-1 is the name of the field in which the values of the entity type on the right hand side

of rel-n are recorded.

retrieval-access is a three-character code constructed as follows:

Position 1: What is the retrieval key?

r retrieval on whole entity value
w retrieval on individual words in a text, name, or term
b retrieval on whole entity value and on individual words
x retrieval on individual words for some entries, on phrases or whole entity values for others
n no retrieval access through index or limit

Position 2: Inclusive searching and truncation (if any)

u inclusive up
d inclusive down
b inclusive up and down
r right truncation
l left truncation
t right and left truncation or masking
z none of these (use only if position 3 is to be specified)

Note: If a search program allows for sequential retrieval through string search (as indicated in the description of the search program), usually all fields for which it is meaningful can be so searched with left and right truncation.

Position 3: How is the retrieval descriptor specified?
(Default: Full character string of authorized value.)

k key derived from full character string (e.g. OCLC)
u enough characters to make unique authorized value
s authorized value or synonym, in full
t authorized value or synonym, enough characters to make unique

For a printed db-impl, the sort sequence is given instead by SORT-KEY 1, SORT-KEY 2, etc.

The following elements are recorded only if needed.

field-name-2 (optional). Often a field is indexed in a combined index, which may have a field-name-like prefix; this prefix is recorded here. If the combined index has no prefix (like a basic index in DIALOG), use "".

field-name-3 (optional). If a field is indexed in a combined index, it is usually possible to restrict retrieval to entries from the field by using a suffix, like /AB. The suffix is usually the field name, and nothing needs to be recorded in that case. If restriction by field is not possible, use "". Sometimes restriction is possible by a subgroup of the fields combined in the index, e.g., title, controlled terms, free descriptors, but not abstract; record the suffix used for such a subgroup, e.g., QF.

LIMIT is recorded if retrieval through limit is possible. If limit is possible only for selected values, use LIMIT-SELECT.

(**VALUES value-1 ... value-n**) is needed if and only if retrieval access is by limit, but only for selected values (LIMIT-SELECT).

RANGE indicates that a search key range (e.g. a date range) can be searched.

NO-DISPLAY is entered only when appropriate; same for SORT.

db-1 **has-index** ((rel-1 ... rel-n db-2) db-3)

db-1 and db-3 are both members of a more comprehensive db. The head-entity-type of db-1 (the database indexed) must agree with the entity-type on the left side of rel-1. db-3 (the index) has as head-entity-type the entity type on the right side of rel-n. db-2 is given only if there are two or more gives-data-on statements for the same relationship chain as in the Excerpta Medica example above.

4.4 Relationships to record information about user support

Statements about user support may apply to a db-gen - support provided by the database producer, to a db-impl - support provided by a search service for the db-impl at hand, or to a search service in general - support for all the db-impl it offers.

db **search-support-overall** rating
legal-ent

db **search-support-material** (document rating)
legal-ent

db-impl **search-support-online** rating
legal-ent

db-impl **search-support-mailbox** rating
legal-ent

db **search-support-phone** (phone-no rating)
legal-ent

4.5 Recording global rules for the selection of databases

Request-property **suggests** (db strength-of-suggestion)

See * for a discussion of this type of statement

4.6 Relationships for recording data about selection criteria

4.6.1 Introduction

Defining database coverage is done in two steps. In step 1 the selection criteria are defined in general; these selection criteria are part of the first component of the knowledge base because the same selection criterion may apply to several databases. Step 2, then, is making the connection between a database and the selection criteria it uses.

The selection criteria (sel) are used in database descriptions

- to define the coverage of a database;
- to rate the coverage and timeliness of a database as a whole or with respect to subsets within a database (e.g., with respect to all English-language documents in a bibliographic database covering other languages too);
- to give information about overlap between databases with respect to a criterion (two databases may show great overlap in one area and very little in another).

4.6.2 Construction of selection criteria

Selection criteria are just like query formulations; they select a subset of entity values (more generally, a subset of sentences). Thus, the construction of selection criteria described in the following is just like the construction of query formulations.

Simple selection criteria are defined in terms of relationships.

sel-number	**is-simple-criterion**	(rel expression)
sel-153	is-simple-criterion	(rel-002 Biomedicine)

where rel-002 is-rel (entity is-about-or-relevant-for-subject)
This defines a simple selection criterion for any type of **entity** (left side of rel-002) by **subject** (right side of rel-002). In order for an entity to be selected, entity is-about-or-relevant-for **Biomedicine** must hold.

The general rules for defining a simple selection criterion are as follows. The rel links entity type 1, from which selection is to be made (left side), with the entity type 2, on which selection is based (right side). Expression gives one or more values of entity type 2. An entity of type 1 meets the selection criterion if it is linked, with rel, to one of the values of entity type 2.

A selection expression may be any of the following:

An entity value, e.g. Biomedicine
A list of the form (LISTED-IN db)
An expression with relational operators

(It was considered to also allow for several values of entity type 2 in a selection expression, as in

sel-213	is-simple-criterion	(rel-002 (VALUES Biomedicine Health-administration))

This would be simpler than using OR, but it would obscure the relationships between selection criteria and make the algorithm for matching query formulations with selection criteria more complex.)

A LISTED-IN expression refers to all entity values in a database. For example, the database ABDA-PHARMA includes a member database of drugs (ABDA-PHARMA-FAM/ZEN) and a member database of organizations (ABDA-PHARMA-HER). An organization is selected for inclusion in ABDA-PHARMA-HER if it produces any substance listed in ABDA-PHARMA-FAM/ZEN. As a selection criterion, this is expressed as follows:

sel-143 is-simple-criterion ((rel-082) (LISTED-IN ABDA-PHARMA-FAM/ZEN))
where rel-082 is-rel (organization produces entity)

Another example expresses the selection criterion for all the entities dealt with in documents contained in MEDLINE:

sel-175 is-simple-criterion (rel-097 (LISTED-IN MEDLINE))

where rel-097 is-rel (entity dealt-with-in document)

As a last example consider PsychIndex, a German index of psychological literature which covers documents authored by psychologists born in a German-speaking country. Their list of psychologists is a database, let's call it VDP (Verzeichnis Deutschsprachiger Psychologen). The selection criterion for PsychIndex, then, is

sel-328 is-simple-criterion (rel-008 (LISTED-IN VDP))

A selection expression may use operators (= being the default):

sel-043 is-simple-criterion (rel-015 (> TODAY - m12))

where rel-015 is-rel (entity entered-in-db date)

sel-043, then, selects all records entered into the database within 12 months from TODAY (TODAY being a system-defined variable containing the current date).

Expressions other than simple entity values appear as Prolog lists.

Selection criteria can be combined with AND and OR to form more complex selection criteria:

sel-number **is-or-criterion** (sel-1 sel-2 ... sel-n)

sel-number **is-and-criterion** (sel-1 sel-2 ... sel-n)

For example, the selection criterion **English language journal articles in Biomedicine** can be expressed as follows:

sel-153 is-simple-criterion (rel-002 Biomedicine)
where rel-002 is-rel (entity is-relevant-for or about subject)

sel-154 is-simple-criterion (rel-034 Journal-article)

where rel-034 is-rel (entity is-of entity-type)

sel-155 is-simple-criterion (rel-016 English)
where rel-016 is-rel (document written-in language)

sel-156 is-and-criterion (sel-153 sel-154 sel-155)

Other combinations can also be defined:

sel-157 is-and-criterion (sel-153 sel-154)

sel-158 is-and-criterion (sel-153 sel-155)

As a final example, consider the definition of a selection criterion for the database VDP mentioned above:

sel-321 is-simple-criterion (rel-034 person)
(entity is-of-type person)

(If **person** is given as the head entity type of VDP, this component is implied and need not be given explicitly)

sel-322 is-simple-criterion (rel-130 Germany)
(person is-born-in Germany)

sel-323 is-simple-criterion (rel-130 Austria)

sel-324 is-simple-criterion (rel-130 Switzerland)

sel-325 is-or-criterion (sel-322 sel-323 sel-324)

sel-326 is-and-criterion (sel-321 sel-325)

Apart from its power in defining selection criteria, the method described here allows for a match of the user's query formulation with the selection criteria defined for the description of databases. In the most usual case, the selection criterion is broader than the query formulation. For example, the search topic **English documents about lung cancer** is formulated as

qf-1 is-simple-qf (rel-002 Lung-cancer)

qf-2 is-simple-qf (rel-016 English)

qf-3 is-and-qf (qf-1 qf-2)

sel-153 (rel-002 Biomedicine) is broader; a database covering sel-153 should contain relevant documents. To be able to select **Lung-cancer** documents, the implementation to be searched must offer retrieval by rel-002 **is-about-or-relevant-for**, and to select **English** documents, it must offer retrieval by rel-016 **Language**. sel-158 (rel-002 Biomedicine AND rel-016 English) is also broader. But one of its component criteria, rel-016 English, is the same as the query formulation component criterion qf-2. Therefore, when searching a database with coverage sel-158, the query formulation can be simplified to just qf-1, and the requirement of searchability by

rel-016 Language can be dropped. There might be a database whose coverage is coextensive with the query topic, such as a bibliography of English language documents on lung cancer. In that case no search in the database is needed; all documents are relevant. Finally, a database may overlap with the query topic in the sense that some, but not all, documents relevant to the query fall within the coverage of the database. Consider

qf-3 English language documents on lung cancer and

sel-157 Biomedical journal articles

A database covering sel-157 should clearly be searched, but it delivers only journal articles, not books and reports. It can be seen from these examples that matching query formulations with selection criteria is not a trivial process, but that the representation chosen here supports it well.

Two databases may have the same selection criterion. More importantly, the building blocks from which compound selection criteria are constructed are shared among databases. For this reason, selection criteria, while introduced to describe a certain database, are defined in the general part of the knowledge base. Their linkage to a specific database is part of the description of that database as discussed in the next section. Since the determination of required database coverage is closely linked to query formulation, this assignment makes sense in terms of the division of labor in the search program.

4.6.3 Application of selection criteria to defining database coverage and overlap

The coverage of a database is described in terms of these selection criteria using the following relationships:

db **covers** (sel proportion completeness accuracy-of-data days)

Proportion is the percentage of the number of entities in the database meeting the sel over the number of all entities in the database, expressed as a percentage or as a rating. Percentage figures are often available or can be quickly estimated from a small sample. Proportion is analogous to precision.

Completeness is the percentage of the number of entities in the database meeting the sel over the total number of entities anywhere meeting the sel expressed as a percentage or a rating. Percentage figures may be available from published studies, but are often hard to obtain. Ratings may be available from reviews or from reference tools describing databases. Completeness is analogous to recall.

Days is the number of days between appearance of an entity or statement and its inclusion in the database.

For each database there is one **covers** statement with a proportion of 100; it defines the coverage of the database as a whole. (Thus there must be a selection criterion that is coextensive with the database coverage.) There can be any number of additional **covers** statements that give information about the coverage of subtopics, specific languages (for example, how well does MEDLINE cover biomedical literature in Russian), forms of materials, etc.; these have a proportion of less than 100. (The selection criteria used in these statements are more specific.) The user wants to know - and the guide should consider - completeness of coverage with respect

to the specific search topic. It is not possible to record completeness in that much detail. But it is possible to enter perhaps 20 **covers** statements for MEDLINE. The guide would then use the statement that best matches the search topic. This provides a better estimate than the completeness for the database as a whole. (In the example, completeness for Russian literature is probably lower than overall completeness.) Accuracy of data and delay may also vary from one portion of a database to another.

db **excludes** sel

Sometimes coverage is described most conveniently by giving a more or less broad area and then listing a few specific exclusions rather than giving a long list of specific areas.

db-1 **overlaps** (sel (db-2 ... db-n) percent date-a date-b)
or
entity-type

This relationship expresses one-sided overlap: What percentage of the entity values (e.g. documents, substances) occurring in db-1 occur also in the union of db-2 ... db-n (i.e., in at least one of the other databases). The one-sided overlap of db-1 with one other database is a special case. Assume db-2 ... db-n have already been searched. The overlap estimates the number of additional documents to be expected in a search of db-1. When using coverage to select databases, the guide would first choose the database with the highest expected return, then the database with the highest return of **additional** documents (which due to overlap may not be the next biggest database). In a more sophisticated version, the guide may combine coverage information with a estimate of recall to be achieved **within** the database to arrive at an effective expected return.

If the first argument on the right-hand side is an entity-type, all entity values of the type are considered. If it is a sel, the sel implies an entity type; only the entity values conforming to the sel are considered in computing overlap. This makes it possible to record such facts as the overlap between MEDLINE and BIOSIS in the genetics literature, in non-English literature, or by any other criterion one might imagine. Since one-sided overlap is not symmetric, two overlap facts should be included for any pair of databases for which the data are available. Overlap of the journal lists of two abstracting services is recorded in the database descriptions for these journal lists and could be used in a pinch as estimate of overlap of the abstracting services themselves.

A date range is included in the **overlap** relation because data are usually available only for certain date ranges. In using these data on generally assumes that they hold for other time periods as well. (The time aspect could also be handled by defining segments for each of the databases and then give overlap data for these segments. This would be more elegant in terms of the relationship definitions but cumbersome in application.)

sel **ranking-by-overall-quality** (db-1 ... db-n)

This relationship allows for the capture of imprecise data. Note that several rankings (from different sources) may be entered.

sel **ranking-by-completeness** (db-1 ... db-n)

sel **ranking-by-accuracy** (db-1 ... db-n)

sel **ranking-by-delay** (db-1 ... db-n)

4.7 Relationships about entity types and relationship types and thesaurus relationships

These relationships are used in knowledge base component 1, which contains a controlled list of the entity types and relationship types found in the databases included in the guide, thesauri, and data about selection criteria (see section 4.6.2).

4.7.1 Relationships about entity types and relationship types

entity-type **ent**

A one-place relation to provide a list of all entity types.

entity-type **is-subtype-of** entity-type

person is-subtype-of legal-ent
journal is-subtype-of serial
serial is-subtype-of document

entity-type **is-component-of** entity-type

city is-component-of address

This relationship is used in searching. For example, if the query requires a database that gives the city where an organization is located, a database that gives the address is found.

rel-number **is-rel** (entity-type-1 relationship-type entity-type-2)

rel-039 is rel (entity has translated-title phrase)
rel-054 is-rel (entity produced-by organization)

or conversely

(entity-type-1 relationship type entity-type-2) **has number** rel-number

Values for entity-type-1 and entity-type-2 must come from the list of entity types. The relationship number is used in statements about databases as explained in Section 4.3.3.

For ease of processing, data about relationship types are stored internally in a format that makes relationship-type a relation in Prolog:

entity-type-1 relationship-type (entity-type-2 rel-number)

entity produced-by (organization rel-054)

rel **inverse-of** rel

The sequence of entity types in a relationship is of great significance in sel and gives-data-on statements. Sometimes, both sequences are needed. For example:

rel-054 is-rel (entity produced-by organization)

rel-082 is-rel (organization produces entity)

Therefore

rel-054 inverse-of rel-082

rel **is-subrel-of** rel

has-translated-title is-subrel-of has-title

Establishes a hierarchy of relationship types to be used in searching.

4.7.2 Thesaurus relationships

The usual relationships between terms and concepts (ST, BT, NT, RT, perhaps with more refinement) serve to represent thesaurus data.

4.8 Additional information

In a complete implementation of the schema it is envisioned that each piece of information (each statement formed using one of the relationship types) can be qualified by a time range for which it is valid, by a time when it was entered, and a source. Furthermore, each statement can be accompanied by a free-form comment.

For relationships with a rather long name it may be useful to introduce an abbreviated name if that would save storage space.

5. Conclusion

The foregoing section did set forth in great detail a conceptual schema for the description of databases in the knowledge base of a database guide. The detail is justified because of the quality of performance in selecting databases is critically dependent on the type of information stored about these databases. This is why our schema allows for recording a large amount of information.

Some principle of our conceptual schema design bear highlighting:

- The recursiveness of the schema. Component databases, thesauri and rule works, and lists of sources are all themselves databases and can be described using the schema. This makes for more elegant descriptions but also puts higher requirements on the search program.

- The use of the entity-relationship approach to the description of database content in conjunction with a thesaurus of entity types and relationship types that are standardized across database descriptions. The inclusion of relationships **is subtype-of**, **is-component-of**, **is-subrelationship-of**, **is-inverse-of** in this thesaurus assists the search program. Queries are also expressed in terms of entity types and relationship types, providing a common basis for

matching. This approach also supports the view that all databases available for one giant database in which queries can be answered by combining data from several relations. The database guide finds the individual databases that hold the relations required.

- The use of relationship chains to define data fields in databases that are not normalized (for example, when the organizational affiliation of an author is given in the record for a document). This makes for a lucid and explicit description of database content that enables the search program to find non-obvious databases (such as MEDLINE when the organizational affiliation of person known to be in the biomedical field is wanted).

- Definition of selection criteria in terms of entity and relationship types based on the recognition that selection criteria are really query formulations that select from the universe into the database. This approach provides great power and flexibility in the definition of selection criteria. The direct parallel with query formulations also allows the system to alert the user when certain specifications should be omitted from the query formulation since they merely duplicate restrictions already imposed in database selection.

- Concise description of database implementation characteristics with respect to retrieval access (especially the indexes available). This can be used to provide specific assistance to users.

A database guide organized along these lines would put a powerful tool in the hands of searchers.

Acknowledgements

The conceptual schema was developed in the framework of a project course at the University of Konstanz in the spring of 1986. The students in the course were a pleasure to work with and they made substantial contributions:

Example

BIOSIS (Biological Abstracts)

BIOSIS produced-by BIOSCIENCES-INFORMATION-CENTER
BIOSIS has-subset BA
BIOSIS has-subset BA-RRM
BIOSIS implemented-as (BIOSIS-DS online data-star $)
BIOSIS implemented-as (BIOSIS-DM online DIMDI $)
BIOSIS implemented-as (BIOSIS-DL online DIALOG $)
BIOSIS-DL has-segment (DL-5 1974 $)
BIOSIS-DL has-segment (DL-55 1969 1973)
BIOSIS-DM has-segment (BA70 1970 1975)
BIOSIS-DM has-segment (BA76 1976 1979)
BIOSIS-DM has-segment (BA80 1980 $)
BIOSIS-DS has-segment (BI77 1970 1977)
BIOSIS-DS has-segment (BIOL 1978 $)
BIOSIS suitable-for (biologists $ 1)
BIOSIS suitable-for (pharmacologists $ 3)
BIOSIS suitable-for (biochemists $ 2)
BIOSIS suitable-for (medical-researchers $ 2)
BIOSIS suitable-for ($ research 1)
BIOSIS suitable-for ($ pure-science 1)
BIOSIS uses-source-direct (books *)
BIOSIS uses-source-indirect BIOSIS-serials-list
BIOSIS uses-source-indirect BIOSIS-proceedings-list
BIOSIS total-entries (document 1980.05 3000000)
BIOSIS is-updated (1980.06 m 23000 * * *)
BIOSIS treats-as-subtype (organism subject)
BIOSIS has-head-entity-type document
BIOSIS gives-data-on ((rel-001) text R-eng)
BIOSIS gives-data-on ((rel-002) term F-eng)
BIOSIS gives-data-on ((rel-002 b) code BIOSIS-Concept-Code)
BIOSIS gives-data-on ((rel-002 rel-094 b) term BIOSIS-Concept-Code)
BIOSIS gives-data-on ((rel-003) code BIOSIS-concept-code)
BIOSIS gives-data-on ((rel-004) phrase R-BIOSIS)
BIOSIS gives-data-on ((rel-008) name R)
BIOSIS gives-data-on ((rel-008 rel-032) name R)
BIOSIS gives-data-on ((rel-008 rel-032 rel-033) address R)
BIOSIS gives-data-on ((rel-015) yymm R)
BIOSIS gives-data-on ((rel-016) term T-eng)
BIOSIS gives-data-on ((rel-017) yyyy R)
BIOSIS-DS gives-data-on ((rel-017) yy R)
BIOSIS gives-data-on ((rel-020) code CODEN-list)
BIOSIS gives-data-on ((rel-020 rel-004) JN R-abbr)
BIOSIS gives-data-on ((rel-034) term F-eng)
BIOSIS gives-data-on ((rel-035) code R)
BIOSIS gives-data-on ((rel-041) term (ABSTRACT))
BIOSIS gives-data-on ((rel-068) string R)
BIOSIS gives-data-on ((rel-106) string R)
BIOSIS gives-data-on ((rel-107) code BIOSIS-Biosystematic-Code)
BIOSIS gives-data-on ((rel-107 rel-075) name BIOSIS-Biosystematic-Code)
BIOSIS-DL records ((rel-001) AB wr "")
BIOSIS-DL records ((rel-002 F-eng) DE wr "")
BIOSIS-DL records ((rel-002 BIOSIS-Concept-Code) CC r $ $)
BIOSIS-DL records ((rel-002 rel-094 b) CN r)
BIOSIS-DL records ((rel-003) CC r)

BIOSIS-DL records ((rel-004) TI wr "")
BIOSIS-DL records ((rel-008) AU rr)
BIOSIS-DL records ((rel-008 rel-032) AA wr)
BIOSIS-DL records ((rel-008 rel-032 rel-033) AA wr)
BIOSIS-DL records ((rel-015) UD r NO-DISPLAY)
BIOSIS-DL records ((rel-016) LA r)
BIOSIS-DL records ((rel-017) JN n)
BIOSIS-DL records ((rel-020) CO r)
BIOSIS-DL records ((rel-020 rel-004) JN r)
BIOSIS-DL records ((rel-034) DE wr "")
BIOSIS-DL records ((rel-035) $ LIMIT)
BIOSIS-DL records ((rel-041) AB r NO-DISPLAY)
BIOSIS-DL records ((rel-068) JN n)
BIOSIS-DL records ((rel-106) JN n)
BIOSIS-DL records ((rel-107) BC rr)
BIOSIS-DL records ((rel-107 rel-075) BN r)
BIOSIS-DM records ((rel-001) AB n)
BIOSIS-DM records ((rel-002 F-eng) UT wt FT)
BIOSIS-DM records ((rel-002 BIOSIS-Concept-Code) CC r)
BIOSIS-DM records ((rel-002 rel-094 b) CT rt)
BIOSIS-DM records ((rel-003) CT rl)
BIOSIS-DM records ((rel-004) TI wt FT)
BIOSIS-DM records ((rel-008) AU rr "SORT?")
BIOSIS-DM records ((rel-008 rel-032) CS n)
BIOSIS-DM records ((rel-008 rel-032 rel-033) CS n)
BIOSIS-DM records ((rel-015) ED r)
BIOSIS-DM records ((rel-016) LA r *)
BIOSIS-DM records ((rel-017) PY r)
BIOSIS-DM records ((rel-017) SO n)
BIOSIS-DM records ((rel-020) CO r)
BIOSIS-DM records ((rel-020 rel-004) JT wt)
BIOSIS-DM records ((rel-034) DE wt FT)
BIOSIS-DM records ((rel-035) ND r)
BIOSIS-DM records ((rel-068) SO n)
BIOSIS-DM records ((rel-106) SO n)
BIOSIS-DM records ((rel-107) BC rt)
BIOSIS-DM records ((rel-107 rel-075) BC rt)
BIOSIS-DS records ((rel-001) AB w *)
BIOSIS-DS records ((rel-002 F-eng) DE wr)
BIOSIS-DS records ((rel-002 BIOSIS-concept-code) PC r)
BIOSIS-DS records ((rel-002 BIOSIS-Concept-Code) SC r)
BIOSIS-DS records ((rel-002 BIOSIS-Concept-Code) TC r)
BIOSIS-DS records ((rel-003) PC r)
BIOSIS-DS records ((rel-004) TI wr)
BIOSIS-DS records ((rel-008) AU w)
BIOSIS-DS records ((rel-017) YR n)
BIOSIS-DS records ((rel-020) CD LIMIT)
BIOSIS-DS records ((rel-020 rel-004) SO n)
BIOSIS-DS records ((rel-034) DE wr)
BIOSIS-DS records ((rel-035) AN n)
BIOSIS-DS records ((rel-068) SO n)
BIOSIS-DS records ((rel-106) SO n)
BIOSIS-DS records ((rel-107) BC r)

Example

IRCS Medical Science

IRCS has-name Elsevier-IRCS
IRCS has-name IRCS_Medical_Science
IRCS produced-by Elsevier-IRCS
IRCS implemented-as (IRCS-DM online DIMDI $)
IRCS implemented-as (IRCS-DS online data-star $)
IRCS-DS has-segment (ER82 1982 $)
IRCS-DS has-segment (IRCS 1981 $)
IRCS has-total-entries (journal-article 1985.02 1960)
IRCS-DM update-total (1985.07 m 75 * * *)
IRCS-DS update-total (* m / 2 * * * *)
IRCS suitable-for (medical-scientists $ 2)
IRCS suitable-for (bio-medicinists $ 1)
IRCS suitable-for (clinical-researchers $ 1)
IRCS suitable-for ($ research 1)
IRCS suitable-for ($ patient-care 3)
IRCS uses-source-indirect ausub
IRCS gives-data-on ((rel-002) term IRCS-categories)
IRCS gives-data-on ((rel-008) name R)
IRCS gives-data-on ((rel-008 rel-032) name R)
IRCS gives-data-on ((rel-008 rel-032 rel-033) address R)
IRCS gives-data-on ((rel-016) term (English))
IRCS gives-data-on ((rel-017) yymmdd R)
IRCS gives-data-on ((rel-017 b) yyyy R)
IRCS gives-data-on ((rel-020) phrase ("IRCS Medical Science"))
IRCS gives-data-on ((rel-038) phrase R-eng)
IRCS gives-data-on ((rel-046) number R)
IRCS gives-data-on ((rel-068) number R)
IRCS gives-data-on ((rel-119) text R-eng)
IRCS gives-data-on ((rel-110) bibl-info R)
IRCS gives-data-on ((rel-110 rel-008) RA ri)
IRCS gives-data-on ((rel-110 rel-004) phrase R-abbr)
IRCS gives-data-on ((rel-110 rel-017) yyyy R)
IRCS gives-data-on ((rel-119) text R-eng)
IRCS gives-data-on ((rel-120) text R-eng)
IRCS gives-data-on ((rel-121) text R-eng)
IRCS gives-data-on ((rel-122) text R-eng)
IRCS-DM records ((rel-002) CT rt SORT)
IRCS-DM records ((rel-002) CT wt FT SORT)
IRCS-DM records ((rel-008) AU r)
IRCS-DM records ((rel-008 rel-032) CS w)
IRCS-DM records ((rel-008 rel-032 rel-033) CS w)
IRCS-DM records ((rel-016) LA r SORT)
IRCS-DM records ((rel-017) PD wt FT)
IRCS-DM records ((rel-017 b) PY r)
IRCS-DM records ((rel-017 b) SO n)
IRCS-DM records ((rel-020) SO n)
IRCS-DM records ((rel-038) TI w FT SORT)
IRCS-DM records ((rel-046) SO n)
IRCS-DM records ((rel-068) SO n)
IRCS-DM records ((rel-110 rel-004) RJ r)
IRCS-DM records ((rel-110 rel-008) RA r)
IRCS-DM records ((rel-110 rel-017) RY r)
IRCS-DM records ((rel-119) IN wt FT)

IRCS-DM records ((rel-120) MM wt FT)
IRCS-DM records ((rel-121) RD wt FT)
IRCS-DM records ((rel-122) NO wt FT)
IRCS-DM records ((rel-123) UC $ $)
IRCS-DS records ((rel-002) DE r "")
IRCS-DS records ((rel-008) AU r "" SORT)
IRCS-DS records ((rel-008 rel-032) IN w "")
IRCS-DS records ((rel-008 rel-032 rel-033) IN w "")
IRCS-DS records ((rel-016) LG r "")
IRCS-DS records ((rel-017 b) SO n)
IRCS-DS records ((rel-020) SO n)
IRCS-DS records ((rel-038) TI w "")
IRCS-DS records ((rel-046) SO n)
IRCS-DS records ((rel-068) SO n)
IRCS-DS records ((rel-110) CR w "")
IRCS-DS records ((rel-119) ID wr "")
IRCS-DS records ((rel-120) ME wr "")
IRCS-DS records ((rel-121) RS wr "")

Example

RTECS (Registry of Toxic Effects)

This example gives only **gives-data-on** statements and statements that are closely related.

RTECS has-head-entity-type substance
RTECS-group-of-compound-list is-of-type thesaurus
Wiswesser-Line-Notation is-of-type rule-work
RTECS-group-of-compound-list has-main-language eng

RTECS gives-data-on ((rel-015) yymm R)
RTECS gives-data-on ((rel-018) name CAS-name-list)
RTECS gives-data-on ((rel-035) code R)
RTECS gives-data-on ((rel-037) name R-all)
RTECS gives-data-on ((rel-040) text R-eng)
RTECS gives-data-on ((rel-040 rel-044) bibl-info R)
RTECS gives-data-on ((rel-040 rel-044 rel-020) code RTECS-CODEN-system)
RTECS gives-data-on ((rel-040 rel-044 rel-020 rel-004) phrase R)
RTECS gives-data-on ((rel-040 rel-044 rel-046) number R)
RTECS gives-data-on ((rel-040 rel-044 rel-068) number R)
RTECS gives-data-on ((rel-040 rel-044 rel-017) yy R)
RTECS gives-data-on ((rel-045) code CAS-registry)
RTECS gives-data-on ((rel-055) term RTECS-group-of-compound-list)
RTECS gives-data-on ((rel-083) weight-number R)
RTECS gives-data-on ((rel-083) weight-number R)
RTECS gives-data-on ((rel-086) number R)
RTECS gives-data-on ((rel-086 rel-044) bibl-info R)
RTECS gives-data-on ((rel-086 rel-044 rel-020) code RTECS-CODEN-system)
RTECS gives-data-on ((rel-086 rel-044 rel-020 rel-004) phrase R)
RTECS gives-data-on ((rel-086 rel-044 rel-020 rel-008) name R)
RTECS gives-data-on ((rel-086 rel-044 rel-020 rel-008 rel-032) name R)
RTECS gives-data-on ((rel-086 rel-044 rel-046) number R)
RTECS gives-data-on ((rel-086 rel-044 rel-017) yy R)
RTECS gives-data-on ((rel-095) explicit-code Wiswesser-Line-Notation)
RTECS gives-data-on ((rel-095) explicit-code RTECS-empirical-formula-rules b)
RTECS gives-data-on ((rel-097) code EPA-TSCA-8E-number)
RTECS gives-data-on ((rel-097) title R)
RTECS gives-data-on ((rel-097 rel-096) text R-eng)
RTECS gives-data-on ((rel-097) bibl-info R)
RTECS gives-data-on ((rel-097 rel-020) code RTECS-CODEN-system)
RTECS gives-data-on ((rel-097 rel-020 rel-004) phrase R)
RTECS gives-data-on ((rel-097 rel-046) number R)
RTECS gives-data-on ((rel-097 rel-068) number R)
RTECS gives-data-on ((rel-097 rel-017) yy R)
RTECS gives-data-on ((rel-097 rel-105) toxic-data-index-string R)
RTECS gives-data-on ((rel-097 rel-105 rel-108) term RTECS-route-of-application-list)
RTECS gives-data-on ((rel-097 rel-105 rel-109) name RTECS-species-list)
RTECS gives-data-on ((rel-097 rel-105 rel-109) name RTECS-species-group-list)
RTECS gives-data-on ((rel-097 rel-105 rel-111) term RTECS-test-type-list)
RTECS gives-data-on ((rel-097 rel-105 rel-112) term RTECS-toxic-effects-list)
RTECS gives-data-on ((rel-097 rel-105 rel-113) term RTECS-toxic-effects-list)
RTECS gives-data-on ((rel-097) bibl-info R b)
RTECS gives-data-on ((rel-097 rel-020) title ("NIOSH CURRENT INTELLIGENCE BULLETIN") b)
RTECS gives-data-on ((rel-097 rel-046) number R b)
RTECS gives-data-on ((rel-097 rel-017) yyyy R b)
RTECS gives-data-on ((rel-097 rel-020) title ("NIOSH MANUAL OF ANALYTICAL METHODS") c)

RTECS gives-data-on ((rel-097 rel-046) number R c)
RTECS gives-data-on ((rel-097 rel-068) number R c)
RTECS gives-data-on ((rel-114) name ("Carcinogenesis Testing Program"))
RTECS gives-data-on ((rel-114 rel-115) acronym (NTP NCI))
RTECS gives-data-on ((rel-116) acronym R)
RTECS gives-data-on ((rel-117) name (TSCA-inventory))
RTECS gives-data-on ((rel-117 rel-118) acronym (EPA))

RTECS-DM gives-data-on ((rel-048) yymmdd R)
RTECS-DM gives-data-on ((rel-036) seqno R)

Appendix 2

Cumulative list of entity types and relationship types

Table of contents

Introduction

This is a list of entity and relationship types that occurred in the 11 databases whose descriptions are given in Appendix 3. There are plans to develop a universal well-structured thesaurus of entity and relationship types starting from this list.

Entity types in alphabetical order

abstract-indicator ent
accession-number ent
address ent
amount ent
application-route ent
aspect ent
author ent
award-type ent
body-part ent
book ent
check-tag ent
chemsubstance ent
city ent
code ent
condition-of-organism ent
contraindication ent
country ent
database ent
date ent
DB-impl ent
disease ent
document ent
document-type ent
drug ent
drug-ingredient ent
drug-preparation ent
drug-type ent
effect ent
entity ent
entity-type ent
enzyme ent
formula ent
geogr-unit ent
grant-number ent
indication ent
ID-code ent
journal ent
journal-article ent
language ent
legal-ent ent
legislation ent
name ent
number ent
number-of-references ent
organ ent
organism ent
organization ent
P-O-Box ent
page-numbers ent
person ent

Entity types

phrase ent
product ent
project ent
serial ent
series ent
standard ent
street ent
subheading ent
subject ent
subject-broad ent
subject-preexploded ent
subject-w-subheading ent
subset ent
substance ent
substance-type ent
term ent
testing-program ent
text ent
text-gen ent
toxic-data-index-string ent
use-or-harm ent
user-comment ent
zipcode ent

Relationships between entity types

book is-subtype-of document
check-tag is-subtype-of subject-broad
chemsubstance is-subtype-of substance
city is-subtype-of geogr-unit
country is-subtype-of geogr-unit
drug is-subtype-of substance
drug-ingredient is-subtype-of drug
drug-preparation is-subtype-of drug
enzyme is-subtype-of chemsubstance
journal is-subtype-of serial
journal-article is-subtype-of document
name is-subtype-of text-gen
organization is-subtype-of legal-ent
person is-subtype-of legal-ent
phrase is-subtype-of text-gen
serial is-subtype-of document
series is-subtype-of document
standard is-subtype-of document
subheading is-subtype-of subject-broad
subject is-subtype-of subject-broad
subject-preexploded is-subtype-of subject
subject-w-subheading is-subtype-of subject
term is-subtype-of text-gen
text is-subtype-of text-gen
text-gen is-subtype-of subject-broad

P-O-Box is-component-of address
city is-component-of address
country is-component-of address
street is-component-of address
zipcode is-component-of address

subject-w-subheading has-components (link subject subheading)

Relationship types by number

rel-001 is-rel (entity has-abstract text)
rel-002 is-rel (entity is-about-or-relevant-for subject)
rel-003 is-rel (entity deals-intensively-with subject)
rel-004 is-rel (entity has-title phrase)
rel-005 is-rel (entity updated-in-DB date)
rel-006 is-rel (entity is-about-or-relevant-for person)
rel-007 is-rel (entity belongs-to subset)
rel-008 is-rel (document authored-by person)
rel-009 is-rel (document edited-by person)
rel-010 is-rel (document illustrated-by person)
rel-011 is-rel (document translated-by person)
rel-012 is-rel (document compiled-by person)
rel-013 is-rel (document emanated-from organization)
rel-014 is-rel (document published-in-country geogr-unit)
rel-015 is-rel (entity entered-in-DB date)
rel-016 is-rel (document written-in language)
rel-017 is-rel (document published-in date)
rel-018 is-rel (entity has-preferred-name name)
rel-019 is-rel (document contains number-of-references)
rel-020 is-rel (document contained-in document)
rel-021 is-rel (entity is-about-or-relevant-for subheading)
rel-022 is-rel (entity is-about-or-relevant-for subject-w-subheading)
rel-023 is-rel (entity is-about-or-relevant-for subject-preexploded)
rel-024 is-rel (entity is-about-or-relevant-for check-tag)
rel-025 is-rel (project carried-out-by organization)
rel-026 is-rel (project supported-by organization)
rel-027 is-rel (project has-principal-investigator person)
rel-028 is-rel (project has-staff person)
rel-029 is-rel (entity has-sale-regulation text)
rel-030 is-rel (entity is-explained-by text)
rel-031 is-rel (substance has-special-additions text)
rel-032 is-rel (person is-affiliated-with organization)
rel-033 is-rel (legal-ent located-at address)
rel-034 is-rel (entity is-of entity-type)
rel-035 is-rel (entity has-DB-accession-no accession-number)
rel-036 is-rel (entity has-DB-impl-accession-no accession-number)
rel-037 is-rel (entity has-nonpreferred-name name)
rel-038 is-rel (entity has-original-title phrase)
rel-039 is-rel (entity has-translated-title phrase)
rel-040 is-rel (substance has-chemical-definition text)
rel-041 is-rel (document has-abstract-indicator abstract-indicator)
rel-042 is-rel (document is-result-of project)
rel-043 is-rel (project has-grant-number grant-number)
rel-044 is-rel (entity has-source document)
rel-045 is-rel (entity has-ID ID-code)
rel-046 is-rel (document has-vol-info number)
rel-047 is-rel (document summary-written-in language)
rel-048 is-rel (entity entered-in-DB-impl date)
rel-049 is-rel (substance has-special-instructions text)

Relationship types by number

rel-050 is-rel (substance has-emp-formula formula)
rel-051 is-rel (substance belongs-to substance-type)
rel-052 is-rel (substance has-title-number code)
rel-053 is-rel (entity has-trade-name name)
rel-054 is-rel (entity produced-by organization)
rel-055 is-rel (substance has-use-or-harm text)
rel-056 is-rel (substance has-application-route application-route)
rel-057 is-rel (substance contains substance)
rel-058 is-rel (substance has-indication condition-of-organism)
rel-059 is-rel (substance has-contraindication condition-of-organism)
rel-060 is-rel (substance interacts-with substance)
rel-061 is-rel (entity has-dosage text)
rel-062 is-rel (substance marketed-since date)
rel-063 is-rel (substance has-property text)
rel-064 not assigned
rel-065 is-rel (substance has-adverse-effects text)
rel-066 is-rel (substance distributed-by organization)
rel-067 is-rel (document has-issue-info number)
rel-068 is-rel (document has-page-numbers page-numbers)
rel-069 is-rel (entity is-about-or-relevant-for substance)
rel-070 is-rel (legal-ent located-at country)
rel-071 is-rel (legal-ent located-at city)
rel-072 is-rel (legal-ent located-at zipcode)
rel-073 is-rel (legal-ent located-at P-O-Box)
rel-074 is-rel (legal-ent located-at street)
rel-075 is-rel (entity has-name name)
rel-076 is-rel (entity has-availability-indicator code)
rel-077 is-rel (drug belongs-to drug-type)
rel-078 is-rel (substance contains text)
rel-079 is-rel (substance has-fragment formula)
rel-080 is-rel (substance contained-in substance)
rel-081 is-rel (entity sold-in geogr-unit)
rel-082 is-rel (organization produces entity)
rel-083 is-rel (substance has-molecular-weight number)
rel-084 is-rel (entity is-about-or-relevant-for organization)
rel-085 is-rel (document published-by organization)
rel-086 is-rel (substance has-aquatic-toxicity-rating-tlm96 number)
rel-087 is-rel (journal has-subdivision-title phrase)
rel-088 is-rel (project located-at country)
rel-089 is-rel (entity started date)
rel-090 is-rel (entity ended date)
rel-091 is-rel (entity is-about-or-relevant-for enzyme)
rel-092 is-rel (project awarded-through award-type)
rel-093 is-rel (entity produced-number-of-publications number)
rel-094 is-rel (subject has-term term)
rel-095 is-rel (subject has-code code)
rel-096 is-rel (document contains-recommendation-about substance)
rel-097 is-rel (entity is-dealt-with-in document)
rel-099 is-rel (entity has-aspect-info aspect)
rel-100 is-rel (entity is-about-or-relevant-for chemsubstance)
rel-101 is-rel (entity is-governed-by standard)
rel-102 is-rel (entity is-about-or-relevant-for geogr-unit)

Relationship types by number

rel-103 not assigned
rel-104 not assigned
rel-105 is-rel (document is-source-of toxic-data-index-string)
rel-106 is-rel (document has-vol-issue-info number)
rel-107 is-rel (entity is-about-or-relevant-for organization)
rel-108 is-rel (toxic-data-index-string contains-info-on route-of-application)
rel-109 is-rel (toxic-data-index-string contains-info-on organism)
rel-110 is-rel (document cites document)
rel-111 is-rel (toxic-data-index-string contains-info-on test-type)
rel-112 is-rel (toxic-data-index-string contains-info-on use-or-harm)
rel-113 is-rel (toxic-data-index-string contains-info-on organ)
rel-114 is-rel (entity tested-by testing-program)
rel-115 is-rel (testing-program executed-by organization)
rel-116 is-rel (entity is-governed-by legislation)
rel-117 is-rel (document is-part-of database)
rel-118 is-rel (database produced-by organization)
rel-119 is-rel (document contains-intro text)
rel-120 is-rel (document contains-materials-and-methods-section text)
rel-121 is-rel (document contains-results-and-discussion-section text)
rel-122 is-rel (document contains-acknowledgements text)
rel-123 is-rel (DB-impl has-special-service user-comment)
rel-124 is-rel (entity has-generic-name name)
rel-125 is-rel (author works-in geogr-unit)
rel-126 is-rel (entity has-text text)

Relationships between relationship types

rel-080 is-inverse-of rel-057
rel-082 is-inverse-of rel-054

has-adverse-effects is-subrel-of has-aspect-info
has-application-route is-subrel-of has-aspect-info
has-contraindication is-subrel-of has-aspect-info
has-dosage is-subrel-of has-aspect-info
has-indication is-subrel-of has-aspect-info
has-property is-subrel-of has-aspect-info
has-special-instructions is-subrel-of has-aspect-info
has-use-or-harm is-subrel-of has-aspect-info
interacts-with is-subrel-of has-aspect-info
has-nonpreferred-name is-subrel-of has-name
has-preferred-name is-subrel-of has-name
contains-intro is-subrel-of has-text
contains-materials-and-methods-section is-subrel-of has-text
contains-results-and-discussion-section is-subrel-of has-text
contains-acknowledgements is-subrel-of has-text
has-original-title is-subrel-of has-title
has-translated-title is-subrel-of has-title

Relationship types in alphabetical order by first entity

(author works-in geogr-unit) has-number rel-125
(database produced-by organization) has-number rel-118
(DB-impl has-special-service user-comment) has-number rel-123
(document authored-by person) has-number rel-008
(document cites document) has-number rel-110
(document compiled-by person) has-number rel-012
(document contained-in document) has-number rel-020
(document contains number-of-references) has-number rel-019
(document contains-acknowledgements text) has-number rel-122
(document contains-intro text) has-number rel-119
(document contains-materials-and-methods-section text) has-number rel-120
(document contains-recommendation-about substance) has-number rel-096
(document contains-results-and-discussion-section text) has-number rel-121
(document edited-by person) has-number rel-009
(document emanated-from organization) has-number rel-013
(document has-abstract-indicator abstract-indicator) has-number rel-041
(document has-issue-info number) has-number rel-067
(document has-page-numbers page-numbers) has-number rel-068
(document has-vol-info number) has-number rel-046
(document has-vol-issue-info number) has-number rel-106
(document illustrated-by person) has-number rel-010
(document is-part-of database) has-number rel-117
(document is-result-of project) has-number rel-042
(document is-source-of toxic-data-index-string) has-number rel-105
(document published-by organization) has-number rel-085
(document published-in date) has-number rel-017
(document published-in-country geogr-unit) has-number rel-014
(document summary-written-in language) has-number rel-047
(document translated-by person) has-number rel-011
(document written-in language) has-number rel-016
(drug belongs-to drug-type) has-number rel-077
(entity belongs-to subset) has-number rel-007
(entity deals-intensively-with subject) has-number rel-003
(entity ended date) has-number rel-090
(entity entered-in-DB date) has-number rel-015
(entity entered-in-DB-impl date) has-number rel-048
(entity has-abstract text) has-number rel-001
(entity has-aspect-info aspect) has-number rel-099
(entity has-availability-indicator code) has-number rel-076
(entity has-DB-accession-no accession-number) has-number rel-035
(entity has-DB-impl-accession-no accession-number) has-number rel-036
(entity has-dosage text) has-number rel-061
(entity has-generic-name name) has-number rel-124
(entity has-ID ID-code) has-number rel-045
(entity has-name name) has-number rel-075
(entity has-nonpreferred-name name) has-number rel-037
(entity has-original-title phrase) has-number rel-038
(entity has-preferred-name name) has-number rel-018
(entity has-sale-regulation text) has-number rel-029

Relationship types alphabetical

(entity has-source document) has-number rel-044
(entity has-text text) has-number rel-126
(entity has-title phrase) has-number rel-004
(entity has-trade-name name) has-number rel-053
(entity has-translated-title phrase) has-number rel-039
(entity is-about-or-relevant-for check-tag) has-number rel-024
(entity is-about-or-relevant-for chemsubstance) has-number rel-100
(entity is-about-or-relevant-for enzyme) has-number rel-091
(entity is-about-or-relevant-for geogr-unit) has-number rel-102
(entity is-about-or-relevant-for organization) has-number rel-084
(entity is-about-or-relevant-for organization) has-number rel-107
(entity is-about-or-relevant-for person) has-number rel-006
(entity is-about-or-relevant-for subheading) has-number rel-021
(entity is-about-or-relevant-for subject) has-number rel-002
(entity is-about-or-relevant-for subject-preexploded) has-number rel-023
(entity is-about-or-relevant-for subject-w-subheading) has-number rel-022
(entity is-about-or-relevant-for substance) has-number rel-069
(entity is-dealt-with-in document) has-number rel-097
(entity is-explained-by text) has-number rel-030
(entity is-governed-by legislation) has-number rel-116
(entity is-governed-by standard) has-number rel-101
(entity is-of entity-type) has-number rel-034
(entity produced-by organization) has-number rel-054
(entity produced-number-of-publications number) has-number rel-093
(entity sold-in geogr-unit) has-number rel-081
(entity started date) has-number rel-089
(entity tested-by testing-program) has-number rel-114
(entity updated-in-DB date) has-number rel-005
(journal has-subdivision-title phrase) has-number rel-087
(legal-ent located-at address) has-number rel-033
(legal-ent located-at city) has-number rel-071
(legal-ent located-at country) has-number rel-070
(legal-ent located-at P-O-Box) has-number rel-073
(legal-ent located-at street) has-number rel-074
(legal-ent located-at zipcode) has-number rel-072
(organization produces entity) has-number rel-082
(person is-affiliated-with organization) has-number rel-032
(project awarded-through award-type) has-number rel-092
(project carried-out-by organization) has-number rel-025
(project has-grant-number grant-number) has-number rel-043
(project has-principal-investigator person) has-number rel-027
(project has-staff person) has-number rel-028
(project located-at country) has-number rel-088
(project supported-by organization) has-number rel-026
(subject has-code code) has-number rel-095
(subject has-term term) has-number rel-094
(substance belongs-to substance-type) has-number rel-051
(substance contained-in substance) has-number rel-080
(substance contains substance) has-number rel-057
(substance contains text) has-number rel-078
(substance distributed-by organization) has-number rel-066
(substance has-adverse-effects text) has-number rel-065

Relationship types alphabetical

(substance has-application-route application-route) has-number rel-056
(substance has-aquatic-toxicity-rating-tlm96 number) has-number rel-086
(substance has-chemical-definition text) has-number rel-040
(substance has-contraindication condition-of-organism) has-number rel-059
(substance has-emp-formula formula) has-number rel-050
(substance has-fragment formula) has-number rel-079
(substance has-indication condition-of-organism) has-number rel-058
(substance has-molecular-weight number) has-number rel-083
(substance has-property text) has-number rel-063
(substance has-special-additions text) has-number rel-031
(substance has-special-instructions text) has-number rel-049
(substance has-title-number code) has-number rel-052
(substance has-use-or-harm text) has-number rel-055
(substance interacts-with substance) has-number rel-060
(substance marketed-since date) has-number rel-062
(testing-program executed-by organization) has-number rel-115
(toxic-data-index-string contains-info-on organ) has-number rel-113
(toxic-data-index-string contains-info-on organism) has-number rel-109
(toxic-data-index-string contains-info-on route-of-application) has-number rel-108
(toxic-data-index-string contains-info-on test-type) has-number rel-111
(toxic-data-index-string contains-info-on use-or-harm) has-number rel-112