IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4425

# AT THE INTERSECTION OF ONTOLOGICAL DESIGN PATTERNS AND (SEMI-)AUTOMATIC DATABASE ANNOTATION

JACQUELINE R. REICH
University of Manchester, Department of Computer Science, Oxford Road,
Manchester M13 9PL, United Kingdom, 0044-161-275-6138, reich@cs.man.ac.uk

## ABSTRACT

*Ontological Design Patterns (ODP) are a technique to improve the design and implementation of 'ontologies'. 'Ontologies' control the language vocabulary of a knowledge domain, related term definitions, and semantic contexts. In combination with Natural Language Processing (NLP) procedures (i.e. text tagging and parsing), 'ontologies' including ODPs were used to extract textual information from SWISS-PROT database entries and to compare text fragments of scientific abstracts. The 'ontologies' were again applied during database annotation while generating and composing comments out of validated and evaluated textual information units. ODPs proved to increase the reusability of 'sub-ontologies', and their flexibility to adapt to different semantic contexts and application tasks.*

## 1 INTRODUCTION

Within this genome-sequencing era, scientists want to exploit the information of genome sequencing results, protein interactions, and gene deleted phenotypes. The molecular biological literature (i.e. abstracts and papers collected in the National Library of Medicine's MEDLINE databases) and molecular biological sequence databases (i.e. SWISS-PROT) are principal computer-based information resources. The automated extraction of text fragments forming task-specific information is a constant challenge regarding the quality, reliability and usefulness of the final output. The goal is to evaluate and validate the results of syntactic and semantic text analyses and to pipe them into in the comment fields of annotated databases.

The (semi-)automatic support of database annotation is an increasingly important task caused by the never ending flux of new molecular biological data, changing interpretations and expanded knowledge. The amount of information relevant for a specific scientific investigation task can be overwhelming. In order to manage and exploit the information, the background knowledge is modelled within 'ontologies' to become non-ambiguous in defined semantic contexts. There are various approaches to the representation of scientific background knowledge discussed in the literature[3,10,14].

Providing 'ontologies' for complex information is challenging. Ontological Design Patterns (ODPs) make the ontological structure and content explicit[19]. ODPs can be traced to reuse and adapt 'ontologies' and 'sub-ontologies' to objectives of information search, analysis and comparison. These engineering design patterns and the terminologies specific for some protein families, such as opsin, rhodopsin or hydrogenase, were represented within the object-oriented programming language Common Lisp Object System (CLOS)[23]. Part of my ODP related research is to understand how the inheritance mechanism of CLOS classes, the definition of attributes, and the application of methods can increase the expressiveness and independence of individual or composed terms forming actual information and knowledge units.

My current application target is to extract information fragments from SWISS-PROT entries and to expand or validate them by analysing referenced MEDLINE abstracts in order to create lists of thematically grouped information bits of final database annotations. I used a scanner to tokenise the texts of the abstracts and SWISS-PROT database entries, which were tagged by a chart-parser. Special semantic features of the texts were recognised referring to background-knowledge represented in hand-made 'ontologies'. The focus on some particular domain knowledge, such as protein families, provided important constraints on the set of concepts, which were integrated and mutually related within a set of 'sub-ontologies'. I applied some simple statistics to provide an overview of larger text-segments, and to get a hint where to look for interesting or exceptional features.

The approach and examples presented here are very much reduced in their size and complexity and restricted to the comparison of database entries and text fragments, where the role of 'ontology' construction and application is most obvious. In spite of the pedagogical examples, I hope that the principles and processes shown or referenced in the literature will be applicable and transferable to larger applications.

## 2 ONTOLOGICAL DESIGN PATTERNS IN THEORY

Within the bioinformatics community, an 'ontology' generally comprises hierarchically organised concepts, relationships, terminological definitions reduced to a set of attributes, and contexts of validity[10,14]. Ontological Design Patterns (ODPs) are a collection of methods for designing and implementing more effective 'ontologies'. The structure and purpose of ODPs are based on 'sub-ontologies' that form a reusable design or a skeleton of a larger 'ontology'. 'Sub-ontologies' are represented by a set of general and specific concepts, their interactions and dependencies. Without the application of ODPs, the interactions of many concepts, relationships, and definitions valid in various contexts can be a multi-dimensional problem difficult to analyse.

Applications of 'ontologies' seem to be infinitely variable, and any library of 'sub-ontologies' will eventually need to expand or change. ODPs help determine and internally represent the detail structure of variable ontological features. Independent '(sub-)ontologies' can easier be ported among different domains and application contexts (i.e. to improve text analyses). ODPs can improve the replacement of one set of domain-dependent concepts with a new set of conceptual interactions, presupposed that 'sub-ontologies' make no different assumptions about their context. Other-

This paper appears in *Managing Information Technology in a Global Economy,* the proceedings of the Information Resources Management Association International Conference. 

wise, ODP based 'sub-ontologies' allow developers to mix and match 'sub-ontologies' from various resources. The drawback is that the application of ODPs requires extensive domain analysis and engineering before benefits can be realised.

The complexity of the design of ODP based 'sub-ontologies' is reduced by the determination of 1) common features that generalise different conceptual alternatives, 2) different alternatives that realise specific features, and 3) the degree of required variability[11]. The ODPs were tested during the process of information extraction out of hydrogenase related SWISS-PROT database entries and referenced abstracts, as shown in the following section.

## 3 FROM DATABASE ENTRIES TO ABSTRACTS

The analysis approach is to identify protein (enzyme) names and thematically related information units in the text of SWISS-PROT database entries (Fig.1) and of MEDLINE abstracts (Fig.5-7).

```
DE PERIPLASMIC [NIFE] HYDROGENASE SMALL
SUBUNIT PRECURSOR (EC 1.18.99.1)
DE (NIFE HYDROGENLYASE SMALL CHAIN)
CC -!- CATALYTIC ACTIVITY: 2 REDUCED FERREDOXIN
       +2H(+)=2 OXIDIZED
CC     FERREDOXIN + H(2)
CC -!- COFACTOR: BINDS TWO 4FE-4S CLUSTERS AND
       ONE 3FE-4S CLUSTER.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A
       SMALL SUBUNIT.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- MISCELLANEOUS: [FE], [NIFE], AND [NIFESE]
       HYDROGENASES APPEAR TO
CC   REPRESENT THREE DISTINCT ENZYMES HAVING
       HYDROGENASE ACTIVITY.
CC -!- SIMILARITY: 30% OVERALL HOMOLOGY
       BETWEEN LARGE AND SMALL
       SUBUNITS OF [NIFE] AND [NIFESE]
HYDROGENASES.
```

Figure 1: Description and comment of a hydrogenase SWISS-PROT entry.

The interesting and informative terms found in the above SWISS-PROT fragment are shown in Figure 2.

| ENZYMES | HYDROGENASE | ACTIVITY CHAIN | FE | 3FE-4S |
|---|---|---|---|---|
| FERREDOXNI PERIPLASMIC PRECURSOR | HYDROGENLYASE | CLUSTERS LARGE SMALL ONES SUBUNIT THREE TWO | NIFE NIFESE | 4FE-4S |

Figure 2: Terms in SWISS-PROT database entry.

*Hydrogenases* catalyse the reversible oxidation of molecular hydrogen and play a vital role in anaerobic metabolism. Metal-containing *hydrogenases* are subdivided into three classes: *Fe* ('*iron* only') *hydrogenases*; *Ni-Fe hydrogenases*; and *Ni-Fe-Se hydrogenases*. Hydrogen oxidation is coupled to the reduction of electron acceptors (such as oxygen, nitrate, sulphate, carbon dioxide and fumarate), whereas proton reduction (hydrogen evolution) is essential in pyruvate fermentation or in the disposal of excess electrons.

The *Ni-Fe hydrogenases*, when isolated, are found to catalyse both hydrogen evolution and uptake, with low-potential multihaem cytochromes, such as cytochrome c3, acting as either electron donors or acceptors, depending on their *oxidation* state. Both *periplasmic* (soluble) and *membrane-bound hydrogenases* are known.

The *Ni-Fe hydrogenases* are *heterodimeric* proteins consisting of *small* (S) and *large* (L) subunits. The *small* subunit contains three *iron-sulphur* clusters (two [4Fe-4S] and one [3Fe-4S]); the *large* subunit contains a nickel ion. *Small* subunits of *membrane-bound Ni-Fe* hydrogenases contain a C-terminal domain of about 40 residues that is absent in *periplasmic* forms.

Figure 3: Expert knowledge (Words in ***italic*** font are shown in Figure 4)

Based on expert knowledge, such as the text in Figure 3, the following term arrangement could be suggested for the terms in Figure 2 (Fig.4). Plain words (i.e. protein) are potential concepts, italic words (i.e. ***periplasmic***) are potential attributes, that may reference other concepts**.**

| | |
|---|---|
| **protein**<br>**enzyme**<br><br>**hydrogenases**<br><br>***periplasmic***<br><br>***membrane-bound*** | **subunits**<br>**large**<br>**[subunits]**<br>***Ni***<br>**small**<br>**[subunits]**<br>***Fe-S*** |
| **hydrogenases**<br>**Fe [hydrogenases]**<br>***iron***<br>**Ni-Fe**<br>**[hydrogenases]**<br>***iron + nickel***<br><br>***heterodimeric***<br>***subunits***<br>**Ni-Fe-Se**<br>**[hydrogenases]**<br>***iron + nickel***<br>***+ selenium*** | **metal**<br>**Fe**<br>***iron***<br>**Ni**<br>***nickel***<br>**Se**<br><br>***selenium***<br>**S**<br>***sulfur***<br>***sulphur*** |

Figure 4: Terms arranged into a simplistic 'ontology'

This term arrangement includes the following intuitive decisions:

1. To keep the example small and simple, the scientific terms FERREDOXIN, and HYDROGENLYASE were not included. The words ACTIVITY, CHAIN, CLUSTERS, ONE, TWO, and THREE belong to the English language and have to be represented in machine-readable dictionaries.
2. The concept 'hydrogenases' could either be sub-grouped by a concept 'metal-containing hydrogenases' comprising the three sub-concepts Fe, Ni-Fe, and Ni-Fe-Se hydrogenases, or the 'metal-containing' feature could be inferred from the sub-concepts, that all refer to a metallic chemical compound.
3. The name of a chemical abbreviation or compound is represented as an attribute. In general, the unfolding of abbreviations could be solved by a look-up in external specific dictionaries. This would also prevent the double coding of orthographic variations, such as 'sulfur' and 'sulphur'.
4. The integration of general English words, such as 'large' and 'small', to specify distinguished concepts indicates the following problem. If texts are scanned for termini techinici (i.e. periplasmic) to gather potential concept names, and lists of stop-words are used to reduce the amount of term-matching then these unspecific terms will not be recognised. Higher term frequencies have to be localised in narrow text fragments to make uncommonly frequent stop-words visible.

After the analysis and exploitation of the information available in SWISS-PROT database entries, and the construction of first 'ontologies', it could be interesting for database annotators to (semi-) automatically investigate the referenced abstracts. This includes 1) to localise statements in abstracts, which are related to information found in the database comment topics and are integrated in the 'ontologies' (Fig. 4), and 2) to check the recognised statements and 'ontologies' for potential contradictions. In the first abstract (Fig. 5), the terms matching the original 'ontology' are shown in italic, and the terms interesting to expand the original 'ontology' are underlined.

"Identification of three classes of ***hydrogenase*** in the genus, Desulfovibrio."

"A comparison of amino-terminal amino acid sequences from the *large* and *small subunits* of *hydrogenases* from Desulfovibrio reveals significant differences. These results, in conjunction with antibody analyses, clearly indicate that the *iron*, *iron + nickel*, and *iron + nickel + selenium* containing *hydrogenases* represent three distinct classes of *hydrogenase* in Desulfovibrio."

Figure 5: Abstract fragment with ***'ontology' terms*** and interesting terms

There is no obvious contradiction between the original 'ontology' and the first abstract, as far as the ontological concepts and attributes are matched in affirmative sentences. The underlined words are potential hooks to compare text fragments in a more detailed way, such as the information about the 'amino-terminal amino acid' sequence being different for 'small' and 'large subunits'.

A further analysis step leads to the comparison of two abstracts, such as the second and third abstract (Fig. 6,7).

"Cloning, characterization, and sequencing of the genes encoding the *large* and *small subunits* of the *periplasmic* [*NiFe*]*hydrogenase* of Desulfovibrio gigas."

"The structural genes for the *large* and *small subunits* of Desulfovibrio gigas *periplasmic* [*NiFe*]*hydrogenase* were identified and isolated by immunological and oligonucleotide screening." ... "Comparison of the amino acid sequence of this *enzyme* with those of two other classes of *hydrogenase* found in Desulfovibrio revealed that the D. gigas *periplasmic hydrogenase* has some homologies to the *periplasmic* [*NiFeSe*]*hydrogenase* of D. baculatus but none to the *periplasmic* [*Fe*]*hydrogenase* of D. vulgaris. ..."

Figure 6: Abstract fragment with *'ontology' terms* and interesting terms

"Crystal structure of the *nickel-iron hydrogenase* from Desulfovibrio gigas."

"... the *heterodimeric Ni-Fe hydrogenase* from Desulfovibrio gigas, the *enzyme* responsible for the metabolism of molecular hydrogen, ... The active site, which appears to contain, besides *nickel*, a second metal ion, is buried in the 60K *subunit*. The 28K *subunit*, which coordinates one [*3Fe-4S*] and two [*4Fe-4S*] clusters,..."

Figure 7: Abstract fragment with *'ontology' terms* and interesting terms

The second abstract (Fig. 6) recommends integrating the reference of each metal-containing hydrogenase to the corresponding Desulfovibrio species. The hierarchical grouping of the genus and species names will form a very simplistic taxonomy including the abbreviated species names (Fig. 8, left side).

| | |
|---|---|
| **Desulfovibrio** | **hydrogenases** |
| **Desulfovibrio baculatus** | **Fe [hydrogenases]** |
| ***D. baculatus*** | ***iron*** |
| **Desulfovibrio gigas** | ***D. vulgaris*** |
| ***D. gigas*** | **Ni-Fe [hydrogenases]** |
| **Desulfovibrio vulgaris** | ***iron + nickel heterodimeric subunits*** |
| ***D. vulgaris*** | ***D. gigas*** |
| | **Ni-Fe-Se [hydrogenases]** |
| | ***iron + nickel + selenium*** |
| | ***D. baculatus*** |

Figure 8: The original simplistic 'ontology' integrating Desulfovibrio species.

Further expansions of the original 'ontology' could include information about 'antibody analyses' in order to match and identify information about 'immunological...screening' in the second abstract (Fig. 6). More difficult, because less direct, would be the identification that '60K' equals a 'large sub-unit', and '28K' equals a 'small sub-unit', such as mentioned in the third abstract (Fig. 7). These simple examples of comparing abstracts and matching information units indicate that a cyclic approach is necessary: Any 'ontology' has to be able to grow and to adapt in a flexible and controlled way.

The next section illustrates the potential power of ODPs applied and integrated within 'ontologies' to keep track of the 'ontology' development, to keep the 'ontology' structure clear, and to exploit the profit of optimal reorganisation of 'sub-ontologies'. It is obvious that (semi-)automatic tools are necessary to handle the enormous complexity of molecular biological knowledge and the challenges caused by information represented in natural language texts. These tools have to be used by human experts who will finally make the decisions concerning the content and design of the growing 'ontology', and the validation of text fragments.

## 4 ONTOLOGICAL DESIGN PATTERNS IN PRACTICE

The design and implementation of complex 'ontologies' remain expensive and error-prone. Much of the cost and effort stems from the continuous rediscovery and reinvention of core concepts and components. ODPs are a promising technology for reusing proven 'ontology' design and implementation in order to reduce the cost and improve the quality of the 'ontology'. They provide modularity, reusability, and extensibility related to 'ontology' development, as shown in the following three sub-sections.

### 4.1 Modularity

ODPs enhance modularity by encapsulating volatile concept implementation details behind stable interfaces. ODP modularity improves 'ontology' quality by localising the impact of design and implementation changes that reduces the effort required to understand and maintain existing 'ontologies'.

For example, the 'Definition Encapsulation ODP' can be applied to define a family of definitions, and to encapsulate each one[21]. Within the 'Interaction Hider ODP', a specific concept encapsulates interacting concepts or instances[21]. The 'Terminological Hierarchy ODP' can be used to compose concepts into part-whole hierarchies concept, or the 'Mask ODP' to represent complete 'sub-ontologies' reduced to a general, summarising concept[21].

*Example: 'Mask ODP'*

In the 'ontology' (Fig. 4), the sub-concept 'Ni-Fe [hydrogenases]' is partially defined by the attribute 'heterodimeric'. A

heterodimer is a protein complex composed of two different polypeptide-chains, that can have the ability to inactivate specific gene regulatory proteins[1]. If the information about a heterodimer gets unfolded then another specific 'ontology' has to be created. Figure 9 only shows a very simplistic version of the concept 'heterodimer'.

| **hydrogenases** | **heterodimer** |
|---|---|
| **Ni-Fe [hydrogenases]** | polypeptide-chain-1 |
| ***iron + nickel*** | *monomer* |
| heterodimeric | polypeptide-chain-2 |
| ***subunits*** | *monomer* |
| ***D. gigas*** | recognised-hybrid-DNA |

Figure 9: Simplistic 'ontology' showing the 'Mask' concept 'heterodimer', which will be referenced by the attribute 'heterodimeric' of the concept 'heterogenases'.

Structuring a larger 'ontology' into 'sub-ontologies' helps reduce complexity minimising dependencies between 'sub-ontologies'. Separated 'sub-ontologies' are more reusable (see next section), and easier to customise. The introduction of the 'Mask' concept 'heterodimer' provides a single, simplified interface and default view to the details of this 'sub-ontology'. The 'heterodimer' concept is relevant for other protein families than hydrogenases, and therefore it may be applied in other contexts than the ones specific for hydrogenases. The 'Mask ODP' can temporarily shield users from 'sub-ontologies', thereby reducing the number of concepts that users (i.e. human beings or software) deal with. This ODP also lets 'sub-ontologies' be varied without affecting the users, and it can eliminate complex or circular conceptual dependencies.

A 'sub-ontology' is analogous to a concept in that both encapsulate something. A concept encapsulates attributes (i.e. features of term definitions) and relationships, and a 'sub-ontology' encapsulates concepts temporarily making them visible or hidden for users.

### 4.2 Reusability

The stable interfaces provided by ODPs enhance reusability by defining generic conceptual components that can be reapplied to create new applications. ODP reusability leverages the domain knowledge and prior effort of experienced developers in order to avoid re-creating and re-validating common solutions to recurring application requirements and ontology design challenges. Reuse of ODP based components can yield improvements in productivity, quality, performance, reliability and interoperability of 'ontologies'.

For example, the 'Expression Composer ODP' designs the same construction process for different concepts, or the 'Unspecific-Term ODP' manages generally unspecific words at fine granularities[21].

*Example: 'Unspecific-Term ODP'*

In the 'ontology' (Fig.4), the sub-concept 'Ni-Fe [hydrogenases]' is partially defined by the attribute 'sub-units'. Supramolecular structures (i.e. enzyme complexes, ribosomes or membranes) are formed by the assembly of pre-formed molecules, which are called 'sub-units'[1]. In the context of hydrogenases, the distinguished sub-units contain either nickel or iron and sulphur. They are characterised as being large or small (Fig.10).

| **hydrogenases** | subunits | **size-adj** |
|---|---|---|
| **Ni-Fe** | **large** | ***small*** |
| **[hydrogenases]** | **[subunits]** | ***large*** |
| **…** | ***Ni*** | ***medium*** |
| subunits | **small** | ***moderate*** |
| **…** | **[subunits]** | |
| | ***Fe-S*** | |

Figure 10: The concept 'sub-units' is referenced by the attribute 'sub-units' of the concept 'heterogenases'.

The 'Unspecific-Term ODP' emphasis the efficient sharing of large numbers of general natural language words, such as 'small' or 'large'. The naive integration of general words into even moderate-sized 'ontologies' of specific terminologies is too expensive. An 'Unspecific Term' is a shared, but independent concept, such as 'size-adj', that can be used in multiple contexts simultaneously. This 'size-adj' concept cannot make assumptions about the context in which it is used, but it declares an interface, which allows users to apply unspecific terms according to external conditions. In general, the interface of an 'Unspecific Term' concept enables sharing, but does not enforce it.

### 4.3 Extensibility

ODPs enhance extensibility by providing explicit hook methods that allow applications to extend its stable interfaces. This is essential to ensure timely customisation of new application services and features.

For example, the 'Terminology Organiser ODP' defines interfaces that create and organise abstract concepts, which will be realised by sub-concepts[21]. Using the 'Add Dynamic Info ODP', information and meaning can be dynamically linked in the form of referenced concepts or attributes[20].

*Example: ' Add Dynamic Info ODP'*

In the 'ontology' (Fig. 8), the sub-concept 'Fe', 'Ni-Fe', and 'Ni-Fe-Se' of hydrogenases are expanded by the corresponding Desulfovibrio species. They are hierarchically grouped within a superficial taxonomy.

The 'Add Dynamic Info ODP' dynamically and incrementally adds additional attributes or information links to a concept, such as the Desulfovibrio species links to the concepts of metallic hydrogenases. This ODP is an alternative to the definition of sub-concepts to extend the original 'ontology' in Figure 4, and prevents the design to use multiple-inheritance or to overwrite basic concepts. The transparency of this ODP allows its recursive application, adding an unlimited number of features. Previously added features can again be withdrawn. This ODP is a mechanism to prevent the explosion of the original 'ontology' when it is threatened by a large number of independent extensions.

## 5 DISCUSSION

The ODPs introduced in Section 4 were implemented in CLOS compiling ontological concepts into CLOS classes, and ontological attributes into CLOS slots. More sophisticated ODP components, such as an active info-adder within the 'Add Dynamic Info ODP', an unspecific-term-manager within the 'Unspecific Term ODP', or a mask-provider within the 'Mask ODP', were realised in appropriate classes and methods to make the various 'sub-ontologies' alive that they can act as the corresponding ODP structure defines or the biological semantic context demands. CLOS or C++ code examples and graphical illustrations for various ODPs can be found in my previous articles [19,20,21].

The use of a particular representation language restricts any ODP based 'sub-ontology' to applications with an identical representation approach. Recently, I could show that it is possible to

write a compiler to transfer an 'ontology' represented in CLOS into an XML based formalism, at least on the syntactic level[18]. It has to be shown in the future, if the (re-)active power provided by the ODP based design for concepts to dynamically operate in various contexts can also be transferred into further formalisms.

During the case study of hand-made 'ontologies' and ODPs, information of abstracts related to the protein families of 'opsin', 'rhodopsin', and 'hydrogenase' was extracted and text fragments were compared. The application of the 'ontologies' and ODPs made it easier to isolate statements relevant for a context specified by a 'sub-ontology'. It was also easier to synthesise the targeted information, such as database annotations. The 'ontologies' made it possible to ignore extraneous and irrelevant information more often, than just context-free lists of specific terms. But the small quantity of analysed abstracts (< 100) and the small, very specific 'ontologies' made a statistical evaluation impossible. Therefore, I do not provide any recall or precision measures. In general, it was obvious that besides having access to machine-readable dictionaries, flexible 'ontologies', a large amount of texts, and sophisticated tools manipulating these different resources, the complexity of the analysed abstracts and the challenge of reliable text comparison do not allow to replace human beings completely. Human beings are still necessary to evaluate and validate analysis results and to finally synthesise database annotations of high quality.

## 6 CONCLUSIONS

The construction of protein-family specific 'ontologies', and the creation of ODPs gave insights into the consequences of applying an object-oriented formalism to represent protein specific knowledge. This contributed to the development of tools to support the (semi-)automatic annotation of biological databases summarising and evaluating texts from various on-line information resources. The case studies showed that the necessary text analyses for the final goal of database annotation impose conflicting demands on text comparison, recognition, evaluation and validation. They ask that the analysed information units selected from the texts be normalising, discriminating, and summarising, as well as accurate.

## ACKNOWLEDGMENTS

This work was done at the Department of Computer Science of the University of Manchester, United Kingdom. Many thanks to Daniel Buchholz for reviewing the manuscript.

## REFERENCES

1. Alberts, B. (et.al.). *Molecular Biology of the Cell*. New York (1994)
2. Andrade, Miguel A. (et.al.) *Bioinformatics*. 15:5: 391-412 (1999)
3. Bader, G. D.; Hogue, Ch. W. *Bioinformatics*. 16:5: 465-477 (2000)
4. Cowie, J.; Lehner, W.. *Communications of the ACM*. 39:1: 80-91 (1996)
5. Eisenhaber, F.; Bork, P. *Bioinformatics*. 15:7/8: 528-535 (1999)
6. Fayyad, U. (et.al.). *Communications of the ACM*. 39:11: 27-34 (1996)
7. Fayyad, U. (et.al.). *Communications of the ACM.* 39:11: 51-57 (1996)
8. Fleischmann, W. (et.al.). *Bioinformatics*. 15:3: 228-233 (1996)
9. Fukuda, K. (et.al.). *Proceedings of PSB'99.* Hawaii (1999)
10. Giudicelli, V.; Lefranc, M.-P. *Bioinformatics*. 15:12: 1047-1054 (1999)
11. Guarino, N.; Welty, Ch. *Proceedings of ECAI-2000* Amsterdam (2000)
12. Guthrie, L. (et.al.). *Communications of the ACM.* 39:1: 63-72 (1996)
13. Humphreys, K. (et.al.). *Proceedings of PSB'00.* Hawaii (2000)
14. Karp, P.. *Bioinformatics*. 16:3: 269-285 (2000)
15. Lewis, D. D. (et.al.). *Communications of the ACM.* 39:1: 92-101 (1996)
16. Medigue, C. (et.al.). *Bioinformatics*. 15:1: 2-15 (1999)
17. Parsons, J.; Wand, Y. *Communications of the ACM*. 40:6: 63-69 (1997)
18. Reich, J. R. *IEEE Int. Symp. Bioinformatic & Biomedial Engineering (BIBE 2000)*, Washington, DC (submitted)
19. Reich, J. R. *Proceedings of DEXA 2000*, London/ Greenwich (2000)
20. Reich, J. R. *Proceedings of AAAI 2000,* Stanford, CA. (2000)
21. Reich, J. R. *Proceedings of GCB'99*: 156-165, Hannover, Germany (1999)
22. Rindfleisch, T. C. (et.al.). *Proceedings of PSB'00*, Hawaii (2000)
23. Steele, G. L., Jr. *Common Lisp the Language*, Digital Press . (1990).
24. Thomas, J. (et.al.). *Proceedings of PSB'00.* Hawaii (2000)
25. Wand, Y.; Wang, R., Y. *Communications of the ACM*. 39:11: 86-95 (1996)