

1997

## Hierarchical Classification as an Aid to Browsing

John R. Rose

*University of South Carolina - Columbia*, [rose@cse.sc.edu](mailto:rose@cse.sc.edu)

Caroline M. Eastman

*University of South Carolina - Columbia*, [eastman@cec.sc.edu](mailto:eastman@cec.sc.edu)

Follow this and additional works at: [https://scholarcommons.sc.edu/csce\\_facpub](https://scholarcommons.sc.edu/csce_facpub)



Part of the [Computer Engineering Commons](#)

---

### Publication Info

Published in *Informatika*, Volume 21, Issue 1, 1997, pages 49-57.

<http://www.informatika.si/>

© 1997 by The Slovene Society Informatika

This Article is brought to you by the Computer Science and Engineering, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact [digres@mailbox.sc.edu](mailto:digres@mailbox.sc.edu).

# Hierarchical Classification as an Aid to Browsing

J. Royce Rose and Caroline M. Eastman  
 Department of Computer Science  
 The University of South Carolina  
 Columbia, South Carolina 29208  
 USA

**Keywords:** unsupervised machine learning, hierarchical classification, browsing

**Edited by:** Rudi Murn

**Received:** July 12, 1996

**Revised:** January 8, 1997

**Accepted:** January 16, 1997

*An approach to browsing large chemical reaction databases is presented. The method that is described builds on earlier work in which unsupervised hierarchical classification was used to extract generalizations of reaction classes from reaction databases for use in reaction knowledge bases. The method described in this paper involves classification based on both semantic and topological features. It supports the creation of deep hierarchies in which succeeding levels represent increasing degrees of abstraction. The creation of a hierarchy allows the user to quickly locate interesting items or classes of items by performing a tree traversal as opposed to sequentially scanning a hit list. In addition, the depth of the resulting hierarchy is determined interactively by the user.*

## 1 Introduction

Browsing is a common information seeking activity and has been extensively studied [2]. Although browsing is not well defined, a variety of definitions have been proposed. What they all have in common is information seeking behavior that involves scanning a (possibly large) number of items looking for something of interest. The items are not restricted in nature; they may be books, grocery items, TV shows, or database records. Browsing is appropriate for searches involving some uncertainty about the goal of the search or about the way to achieve the goal (or both).

Several broad classes of database browsing requests can be identified:

1. Items related to X. X is a known or hypothetical item. If X is a known item, it might or might not be in the database. Items might be related to X because they are similar to X or for some other reason. In a chemical reaction database, a user might request reactions similar to a known reaction; similarity might be determined on the basis of the end product or the reaction conditions.
2. Items characterized by P. P is a set of properties. In a chemical reaction database, possible classes of properties include reaction conditions, i.e., temperature, solvent, catalyst, and pressure, topological changes such as ring closure/opening, and general mechanism such as the base catalyzed nucleophilic mechanism.
3. Items of interest. This is a vague and ill-specified request. However, people sometimes browse with exactly this kind of vague goal in mind. Such searches might be facilitated by knowledge discovery systems. Such a system might be used in a chemical reaction database to look for interesting and previously unidentified groups of reactions.
4. Kinds of items in the database. A user might be interested in finding out what kinds of information are in the database. This form of exploration is facilitated by a classification of the contents of the database. Although this could be done manually, an automatic classification is both more convenient and potentially more flexible. It makes it practical to create several classifications based on differ-

ent dimensions.

Imposing a hierarchical classification, in which succeeding levels represent increasing degrees of abstraction, on either the entire database or the results of a user query can be used to support these four broad classes of browsing requests. The creation of a hierarchical classification on a hit list allows the user to examine the hit list by performing a tree-traversal. This makes it possible to rapidly evaluate the contents of the hit list and quickly locate those items or sets of items the user is searching for or to determine that they are not present. Browsing by traversing such a hierarchy is equivalent to being able to query by similarity. We have chosen to evaluate this approach to browsing in large chemical reaction databases.

## 2 The Domain

Chemistry is the science that among other things deals with the transformations that substances undergo. Two key problems in this field are reaction prediction and synthesis design. Reaction prediction addresses the question of what chemical reaction or reactions will take place with a given starting material under particular conditions. In the case of synthesis design, the chemist has a target compound in mind. The question here is what should be used as starting material and what reaction or series of reactions should be used in order to transform the starting material into the desired target compound.

Reaction prediction and synthesis design require the chemist to have a very good understanding of the types of reactions that may possibly occur with a given set of materials and the influence that reaction conditions have. Where does the chemist get this information? Historically, chemists have learned about chemistry by reasoning from individual examples and by inducing generalizations from sets of related reactions. The chemist may be able to accurately predict the resulting transformation on a set of starting materials if these materials and the reaction conditions are similar to a known reaction. On the other hand, this prediction may also be made possible by an understanding of the underlying chemical processes. This deep understanding can be derived by generalizing from a set of related reactions.

Both inductive generalization and reasoning from individual examples are predicated on the chemist having access to an appropriate collection of reactions. For this reason, chemistry has always been a field in which databases have been compiled. Thus chemistry databases have existed long before the advent of the modern digital computer. In earlier times, these databases took the form of multi-volume compilations much like very large cookbooks. Today the field of chemistry is well supported by computerized databases [7,22]. These databases provide access to information about the scientific literature, chemistry hand books, patent information, business and industry data, chemical substance information, and reaction information. Textual, structural, and factual information is supported. In recent times, databases with more than one million reactions have been compiled [1]. Other reaction databases are growing by as much as 60,000 reactions per year [17].

## 3 The Problem

Chemistry is a field in which the amount of information available has consistently exceeded the capability of database technology. The explosive growth of reaction databases brings its own set of problems. One of the most pressing problems is not how the data is stored but how the user navigates through such a vast amount of information. This is usually not a problem if the database happens to contain the particular piece of information that the chemist is searching for. However, if this information is not contained in the database and the user must search for similar or related data, then current technology does not provide an adequate solution. Query methods that were adequate for reaction databases comprising tens of thousands of reactions are woefully inadequate when the database grows by one or two orders of magnitude.

An important aspect of the problem that users have with such databases relates to finding a good match between the generality/specificity of their queries and the contents of the database. An optimal match results in a hit list containing only that portion of the database the user is actually interested in. Even in very large reaction databases it may be the case that very little of the chem-

istry that the user is interested in is contained by the database. In this case, the user may have to start with a very general query in order to select the examples representing that chemistry. On the other end of the spectrum, the database may contain a rich complement of reactions, perhaps even the actual example the user is interested in. Here, the user will want to restrict the query to focus on the most relevant reaction or set of reactions.

Typically, the user scans the resulting hit list and then modifies the query in order to better target the relevant portion of the database. This may involve submitting a modified query to the entire database or just to the portion contained in the hit list. This type of query modification is both time consuming and wasteful of resources. One of the more tedious aspects occurs when the user must try to extract a summary of the hit list in order to decide how to modify the query. Quite often this is done by glancing at the first few entries and then modifying the query to exclude the kinds of entries in the hit list that the user does not find relevant.

This process of iterative query modification and hit list summarization results in an incomplete *ad hoc* hierarchical classification. Recognition of this fact leads us to propose hierarchical classification based on unsupervised learning as an efficient method for hit list processing in databases of organic reactions. However, since this problem is very general, we expect that many of the lessons learned will be applicable to other domains in which very large databases of complex objects are used.

## 4 Classification Methodology

The approach to hierarchical classification that we have taken is based on both semantic and topological features. It builds on the previous work of Rose and Gasteiger [19,20] which in turn was based on an earlier scheme that primarily considered topological features [14]. It supports the creation of deep hierarchies in which succeeding levels represent increasing degrees of abstraction. Our initial efforts have focussed on classifying the retrieved set (hit list) and not the entire database. We believe that providing the hit list with a hierarchical structure is more related to the needs of the user than would be reorganizing the entire

database. Another reason for not restructuring an entire database at the very beginning is that such an approach would demand extremely close cooperation with a database provider. However, we expect that the experience we gain from structuring hit lists will be valuable for later work involving entire databases.

### 4.1 The HORACE Algorithm

The HORACE hierarchical classification algorithm was developed for classifying and generalizing sets of chemical reactions. The primary motivation for this earlier work was the extraction of generalized reaction descriptions for use in chemical knowledge bases to support synthesis design and reaction prediction systems. Consequently, the hierarchies that are produced are created with the specific goal of producing reaction class descriptions with the degree of abstraction appropriate for a synthesis design or reaction prediction knowledge base. The resulting hierarchy is simply a means to an end. This algorithm for which a detailed description has already been published[19] is shown schematically in Figure 1.

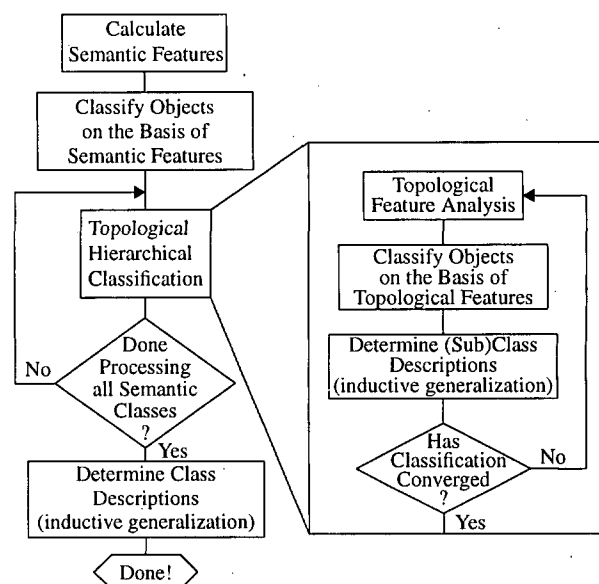


Figure 1: Hierarchical classification algorithm combining semantic and topological metrics.

The algorithm starts by calculating the semantics of the objects being classified. The methods for performing this characterization build on empirical methods developed by the EROS group

during the last 15 years [8, 9, 10, 11, 12, 15]. These are used to characterize the electronic and energy effects operative at the atoms and bonds of the reaction center. Classification at this level then is based on the comparison of corresponding atoms and bonds of the reaction centers of the reactions with respect to the dimensions defined by these parameters.

During the topological phase of hierarchical classification, the reactions are analyzed for topological features to support classification. HORACE uses a list of 114 features which are essentially chemical subgraphs recognized by chemists as functional groups. This set of 114 target features is stored in an external file which can easily be modified by adding or removing features. At this level, the classification of reactions involves the comparison of their complements of topological features. The precise details of HORACE's semantic and topological classification can be found in Rose and Gasteiger[19].

A hallmark of this approach to classification is the alternation between phases of classification and generalization and the way in which semantic and topological classification is combined. A key feature of this algorithm is the manner in which it combines structural and semantic classification approaches. It does not simply compose the two classification methods. Rather, it propagates constraints from the semantic phase of classification into the topological phase. This is done by first computing the semantic classification and then creating a topologically-based hierarchy on each of the resulting clusters (Figure 2). Since the topological algorithm is processing only reactions from one semantic cluster at a time, it cannot mistakenly combine reactions from separate semantic clusters that might appear to be topologically similar. The semantic features in the case of chemical reactions consist of descriptions of chemical structure in terms of electronic and energy parameters. These describe the meaning of the structure and make it possible to create chemically valid equivalence classes of reactions. The semantic classification is extended by alternating phases of topological classification and generalization of both semantic and topological descriptions. After the topological classification stabilizes, a final generalization based on the initial semantic classification is performed.

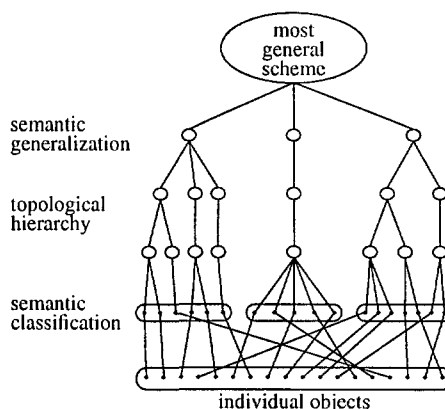


Figure 2: Stylized classification tree.

As can be seen in Figure 2, the topological hierarchical classification actually expands the classification tree in between the semantic classification and the final semantic generalization level. In a given hierarchy, each level represents a different degree of abstraction. The original objects being classified are at the lowest level. These items are then classified on the basis of similarity. The next layer consists of generalizations of the classes formed by classification of these items. Each level in the hierarchy is an abstraction of the level below it. The goal is to provide class summaries which are stored at the next highest level of abstraction in the hierarchy. The topmost item in the hierarchy summarizes all of the objects in the tree and is therefore the most general description.

## 4.2 The Modified HORACE Algorithm

In order to derive substantial benefit from giving hierarchical order to data, the resulting classification trees should strike a balance between depth and breadth. For this reason, one important goal in the design of the classification algorithm was to produce classification hierarchies expressing a large range of abstraction. This requirement motivated the design of a classification algorithm combining both phases of semantic and topological classification. A classification based on semantic features makes it possible to recognize similarity between objects that may be topologically dissimilar. On the other end of the spectrum, consideration of topological features makes it possible to refine a classification by extending it

in the direction of greater specificity in a manner that is intuitive to the chemist.

Notice that the hierarchy shown in Figure 2 is not particularly deep. Typically, HORACE hierarchies have the number of levels shown here. Occasionally, however, hierarchies that are shallower or deeper by one level are produced. This is a result of the data driven nature of the algorithm. If the reactions in a given semantic cluster are either topologically very similar or very dissimilar then only a single level of topological classification will be produced [20]. Clearly, such shallow hierarchies are inadequate for supporting the browsing of large numbers of reactions. Consider the case where the hit list contains several hundred reactions. A hierarchy of only a four or five levels lacks balance between breadth and depth. The resulting hierarchy would look more like a fat bush than a tree and would do little to reduce the information overload placed on the user.

The relative shallowness of the hierarchies produced by HORACE has been overcome by modifying the algorithm to increase the number of levels produced on the basis of semantic classification. This is done by varying the distance threshold which is used to determine cluster membership. The user supplies a starting threshold value and all intervening threshold values interactively so that a well-proportioned hierarchical classification tree, from the perspective of the user, results. Although the computed distances between reactions are normalized by the number of atoms and bonds in the reaction centers, selecting an appropriate threshold will depend on the nature of the reactions under consideration. If the reactions are quite similar, then a very low distance threshold will be required to split the clusters of one level into significantly smaller clusters in a deeper level. The threshold defines the upper distance limit allowable for a reaction to still be considered as matching the elements of a cluster. Lowering the threshold corresponds to requiring a closer degree of similarity. Consequently, the depth of the hierarchy is determined by the user interactively.

Once the size of a semantically based cluster drops below a user-specified size, it is no longer considered for further semantic classification. It is then automatically extended by consideration of topological features using the topological portion of the HORACE algorithm. Recall that each

internal node of the hierarchy contains a description which summarizes the subhierarchy which extends underneath it. Such summaries are particularly helpful in the case of topologically based clusters since the resulting descriptions highlight the structural similarity among the items comprising the subhierarchy.

## 5 A Browsing Example

Evaluation of the modified HORACE algorithm for supporting browsing is being carried out on a subset of the ChemInform-RX reaction database[17]. This set, containing approximately 115,000 reactions, corresponds to the reactions compiled in the database during 1991 and 1992. This data set is being accessed directly without going through a database system.

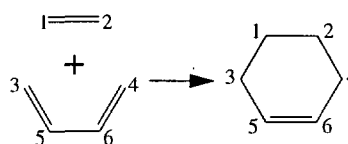


Figure 3: Diels-Alder reaction center.

The transformation shown in Figure 3, which chemists will recognize as the reaction center of the Diels-Alder reaction, was used as a query. The data set was then searched directly using a program written locally. This generated a list of 343 reactions to be treated as a hit list in a simulated reaction database query.

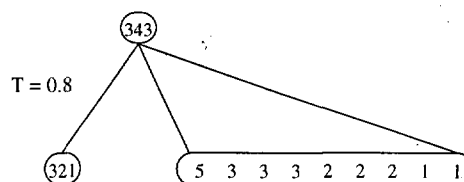


Figure 4: Level 1 of the hierarchical classification.

In Figure 4 we see the first browsing step with the creation of the first and highest level of the hierarchical classification. The user has selected a distance threshold of 0.8 which has partitioned the original 343 reactions into 10 clusters, of which the largest contains 321 reactions. The clusters at each level contain reactions that are mutually dissimilar to those of other clusters at

the same level with respect to the user specified distance threshold. Thus, the twenty-two reactions that are contained in the nine smaller clusters can be interpreted as those reactions most unlike the remaining 321 reactions in the single large cluster. Although a single oval-shaped node has been used to depict the nine smaller clusters in Figure 4 primarily in order to reduce clutter and to make the figure more readable, this depiction also conveys their dissimilarity from the large cluster of 321 reactions. The user that is interested in outliers need only examine these reactions without ever having to scan through the vast bulk that resulted from the simulated query.

As mentioned earlier, once the size of a cluster falls below a user-specified size, it is automatically processed by the topological portion of the HORACE algorithm. The other side of the coin is that these smaller clusters will no longer take part in the refinement of the hierarchy that is based on user selected thresholds. In this particular case, the nine smaller clusters are either so small or similar within a cluster that no further subhierarchy is created on the basis of topological features. However, each cluster is generalized to produce a description which summarizes the cluster content. Thus, the user may choose to look at the generalizations of such small clusters before deciding whether or not to look at the individual reactions. The cluster description shown in Figure 5 is for the cluster containing five reactions from Figure 4. In this figure, the label **R1** denotes the generalization of hydrogen and  $C_{sp^3}$  atoms.

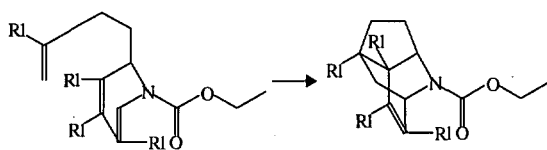


Figure 5: Generalization of the cluster containing five reactions in level 1.

The extension of the classification hierarchy that results from the user having selected a distance threshold of 0.55 followed by a threshold of 0.5 is shown in Figure 6. In the bottom-most level, two large clusters have been produced in addition to 16 smaller clusters. The 16 smaller clusters comprise only 44 of the 305 reactions on this level and in the case of clusters which are not sin-

gletons, the user may initially examine the generalized cluster description before deciding whether or not to look at the individual reactions.

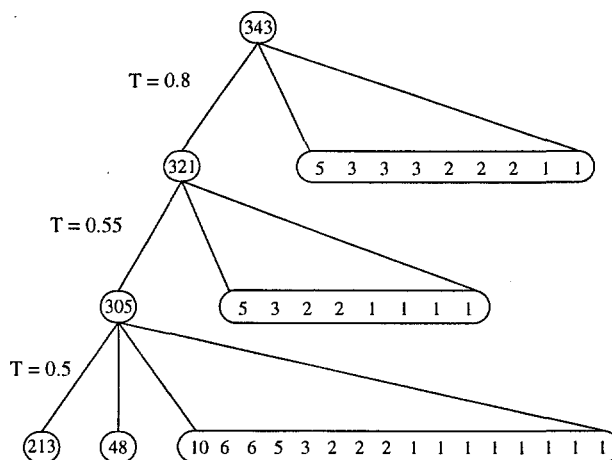


Figure 6: Levels 1-3 of the hierarchical classification.

Figure 7 shows the last level of semantically motivated hierarchical classification construction requested by the user. A threshold of 0.46 was specified by the user to create this level. In noting that the preceding level was created with a threshold of 0.47, we perceive that a critical boundary has been crossed that has resulted in the fragmentation of the cluster containing 202 reactions into 27 clusters, all of which are considerably smaller. It may be reasonable at this juncture for the user to re-specify the cluster size threshold that the system uses to determine when to automatically extend hierarchies with the creation of topologically motivated levels in order to further process the larger remaining clusters. Doing so would, for example, extend the hierarchy rooted at the larger cluster of 59 reactions by the topological-based hierarchy shown in Figure 8. We see that within the span of five user-selected levels the initial monolithic hit list has been systematically reduced to clusters that a user would find much more manageable than the imposing initial set of 343 reactions.

## 6 Related Work

Clustering has been extensively studied across a wide variety of disciplines, and a large number of clustering algorithms have been developed. Many

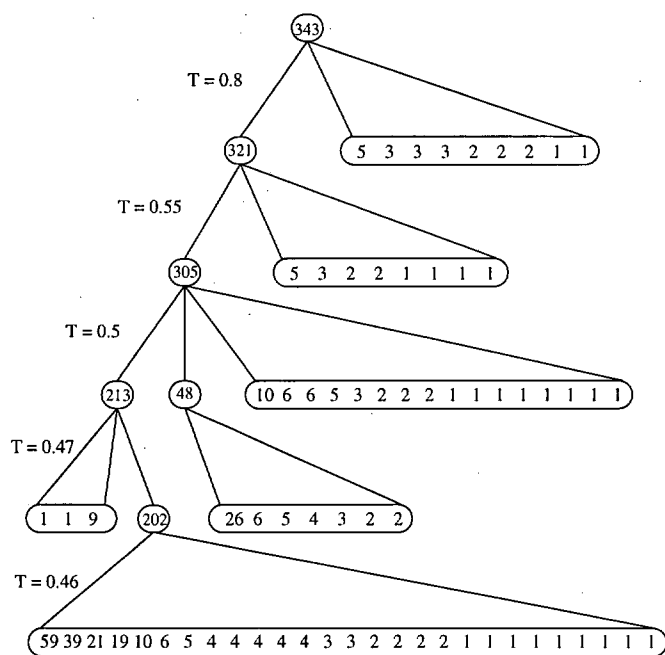


Figure 7: Levels 1-5 of the hierarchical classification.

of these algorithms create hierarchies of clusters, but this is almost always done by splitting or joining existing clusters by varying a cutoff parameter. The algorithm used here is distinctive both in its use of semantic and syntactic information during different phases of the creation of the hierarchy and its use of different, i.e., successively more abstract, information in the creation of each major level in the topological phase. Cutting, Karger, and Pedersen [4] describe a hierarchical approach used in document retrieval applications. However, their clustering algorithm is statistical rather than semantic; it is based on the computation of keyword vector similarity.

Clustering is only one of the techniques that has been used in analyzing and managing chemical information. A complete survey of all of the different approaches proposed or implemented is not feasible here. An overview of storage and processing of chemical structure information is provided by Lipscomb, Lynch, and Willet [16]. They address problems in representation, indexing, and searching in both structure and reaction databases, including similarity based matching and clustering.

## 7 Future Work

The browsing system described in this paper bases its classification purely on topological and physicochemical attributes. The set of 114 topological features used for classification was derived from a collection of functional group structures used by the SYNCHEM synthesis design system [13]. The structures in this subset have not been rigorously evaluated for their appropriateness as classification features. It is expected that some of them could be discarded without negatively affecting classification accuracy. At present, only the physicochemical features sigma and pi electronegativity along with resonance stabilization parameters are used. Additional physicochemical attributes must be evaluated for their classification utility. One area in which the current system is completely lacking is in the use of reaction conditions as classification criteria. Although reaction conditions by themselves can not support the fine degree of classification possible with topological and physicochemical attributes, they are important and must be taken into con-

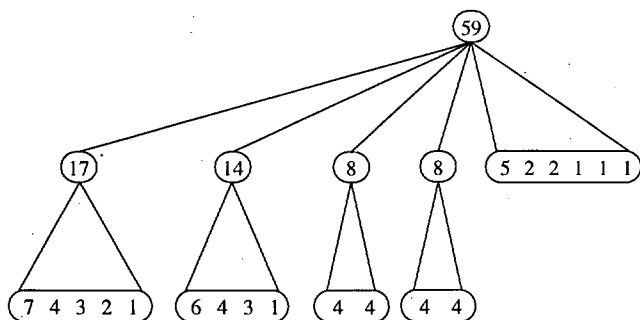


Figure 8: Topological-based hierarchical classification of the level 5 cluster of 59 reactions.



sideration. Additionally, stereo-chemistry is not presently taken into account.

## 8 Conclusion

Hierarchical restructuring allows the user to quickly evaluate the results of a query and to locate interesting items and classes of items. This is accomplished by performing a tree traversal rather than a sequential perusal of a hit list or a series of *ad hoc* query refinements that is normally required for nonhierarchical approaches. More general classes may be examined by moving up the hierarchy. Conversely, more specific classes may be examined by moving down the hierarchy. In contrast, sibling nodes in the hierarchy represent related classes of approximately the same degree of abstraction. In very large databases where classical querying methods are increasingly inadequate such as chemical reaction databases, such a browsing method is required in order to manage the flood of information with which the user is confronted.

There is a long history of interest in intelligent systems to facilitate chemical information processing beginning in the late 1960's. Much of this work has focussed on the development of knowledge-based systems for reaction prediction and synthesis design [3, 5, 6, 13, 18, 21]. The problems of synthesis design and reaction prediction are much more difficult than was thought when research in this field began. Consequently, intelligent systems developed to address these problems have met with limited success. This has been due in large part to the difficulty experienced in compiling adequate knowledge bases. The research that we propose could be adapted to assist in the compilation of chemical reaction knowledge bases since it is essentially a data-mining tool.

## Acknowledgments

The authors would like to thank Prof. Johann Gasteiger for making the empirical methods developed by the EROS group for calculating physicochemical parameters available to us. We would also like to thank Dr. René DePlanque of Fachinformationszentrum Chemie, Berlin, for providing us with the reaction data set used in evaluating the hierarchical classification algorithm.

## References

- [1] J. E. Blake and R. C. Dana, *CASREACT: More than a Million Reactions*. ACS Jour. Chemical Information and Computer Science 1990, 30, 394-399.
- [2] S. Chang and D. E. Rice, *Browsing: a multidimensional framework*. Annual Review of Information Science and Technology 1993, 28, 231-276.
- [3] E. J. Corey and W. T. Wipke, *Computer-assisted Design of Complex Organic Synthesis*. Science 1969, 166, 178-192.
- [4] D. R. Cutting, D. R. Karger, and J. O. Pedersen, *Constant interaction-time scatter/browsing of very large document collections*. In proceedings of the Sixteenth Annual International ACM SIG.R Conference on Research and Development in Information Retrieval. June 27-July 1, 1993, Pittsburgh, PA. pp126-134.
- [5] P. A. D. De Maine and M. M. De Maine, *Computer aids for chemists*. Analytica Chimica Acta 1990, 235, 7-26.
- [6] J. Dugundji and I. Ugi, *An algebraic model of constitutional chemistry as a basis for chemical computer programs*. Top. Curr. Chem. 1973, 39, 19-64.
- [7] J. Garnett and V. Calderhead, *Chemistry. In manual of Online Search Strategies, 2nd Edition*. C. J. Armstrong and J. A. Large, eds. G. K. Hall & Co. New York 1992 pp128-157.
- [8] J. Gasteiger and M. G. Hutchings, *Quantification of Effective Polarisability. Applications to Studies of X-Ray Photoelectron Spectroscopy and Alkylamine Protonation*. J. Chem. Soc. Perkin, 1984, 2, 559-564.
- [9] J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, M. Saller and K. Yuki, *A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design*. Topics Cur. Chem. 1987, 137, 19-73.

- [10] J. Gasteiger and M. Marsili, *Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges*. Tetrahedron, 1980, 36, 3219-3228.
- [11] J. Gasteiger, M. Marsili, M. G. Hutchings, H. Saller, P. Löw, P. Röse and K. Rafeiner, *Models for the Representation of Knowledge about Chemical Reactions*. ACS Jour. Chemical Information and Computer Science 1990, 30, 467-476.
- [12] J. Gasteiger and H. Saller, *Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept*. Angew. Chem. 1985, 97, 699-701. Angew. Chem. Ed. Engl. 1985, 24, 687-689
- [13] H. Gelernter, G. A. Miller, D. L. Larsen and D. J. Berndt, *Realization of a large expert problem-solving system: SYNCHEM2, a case study*. IEEE 1984 Proceedings of the First Conference on Artificial Intelligence Applications. IEEE Computer Society Press: Silver Spring, MD, 1984.
- [14] H. Gelernter, J. R. Rose and C. Chen, *Building and Refining a Knowledge Base for Synthetic Organic Chemistry via the Methodology of Inductive and Deductive Machine Learning*. ACS Jour. Chemical Information and Computer Science, (30):492-504, 1990.
- [15] M. G. Hutchings and J. Gasteiger, *Residual Electronegativity - An Empirical Quantification of Polar Influences and its Application to the Proton Affinity of Amines*. Tetrahedron Lett. 1983, 24, 2541-2544.
- [16] K. J. Lipscomb, M. Lynch and P. Willet, *Chemical structure processing*. Annual Review of Information Science and Technology 1989, 24, 189-238.
- [17] A. Parlow, C. Weiske and J. Gasteiger, *ChemInform - An Integrated Information System on Chemical Reactions*. ACS Jour. Chemical Information and Computer Science 1990, 30, 400-402.
- [18] P. Röse and J. Gasteiger, *EROS 6.0, a Knowledge Based System for Reaction Prediction - Application to the Regioselectivity of the Diels-Alder Reaction*. In Software-Development in Chemistry. Editor: J. Gasteiger. Springer-Verlag, Heidelberg, 1990.
- [19] J. R. Rose and J. Gasteiger, *HORACE: An Automatic System for the Hierarchical Classification of Chemical Reactions*. ACS Jour. Chemical Information and Computer Science, 1994, 34, 74-90.
- [20] J. R. Rose and J. Gasteiger, *Hierarchical Classification as an Aid to Database and Hit-List Browsing*. Third International Conference on Information and Knowledge Management Conference Proceedings, Gaithersburg, Maryland, pp408-414, 1994.
- [21] T. D. Salatin and W. L. Jorgensen, *Computer-assisted mechanistic evaluation of organic reactions*. Journal of Organic Chemistry, 45 (11), 1980.
- [22] H. Schulz and U. Georgy, *From CA to CAS online*. In Databases in Chemistry, 2nd Edition. Springer Verlag, Berlin 1994.