A MOLECULAR CONCEPT OF MANAGING DATA

Christoph Schommer
Department of Computer Science, University Luxembourg, 6 Coudenhove-Kalergi, 1359 Luxembourg, Luxembourg

Keywords: Artificial life, Relational data management, Pattern recognition, Bio-inspired computing.

Abstract: The following (position) paper follows the concept of the field of Artificial Life and argues that the (relational) management of data can be understood as a chemical model. Whereas each data itself is consistent with atomic entities, each combination of data corresponds to a (artificial) molecular structure. For example, an attribute D inside a relational system can be represented by a nucleus α_D sharing a *cloud of values*, which consists of so-called *valectrons* (the values for the column D). By using reaction rules like the selection of tuples or projection of attributes, a retrieve of molecules can be achieved quite easily. Advantages of the chemical model are no data types, a fast data access, and the associative nature of the molecules: this automatically supports a direct identification of patterns in the sense of data mining. A disadvantage is the need for restructuring that must eventually be done, because the incoming data stream is allowed to influence the chemical model. With this position paper, we present our basic concept.

1 INTRODUCTION

Since more than 30 years, relational database systems are successful, being the most important data management system worldwide. Based on the theory on sets, a relational database system takes advantage from the concepts of the relational algebra, which has led – among other functionalities – to today's standard query language SQL. Although in recent past, some alternative database architectures have been developed (object-oriented, object-relational, and XML databases, etc.), they have never received the desired breakthrough. And although the relational systems suffer from an efficient management (often, the more complex the system is the more time and capacity is needed to guarantee the data consistency), the relational architecture still proves reliability, consistency, and precision.

With this position paper, we foster on a completely different approach of data management and try to figure out that data is unlike a data value inside a well-structured environment but even more a fluid (dynamic) and molecular concern. With respect to the natural example, we understand data as an atomic structure and combinations of data as a molecule – invoked on the field of *Artificial chemistry* (Dittrich et al., 2001). Commonly, *Artificial chemistry* is understood as a theoretical model following the natural example, which is used to simulate types of systems in the spirit of chemical reactions (Leach, 2001). It originates in the field of *Artificial Life* (Kelemen and Sosík, 2001) and has proven to be a manifold and powerful pathway of modeling (Skusa et al., 2000), (Ziegler and Banzhaf, 2001), (Schommer, 2009).

In general, an artificial chemistry is defined as a triple (M, R, A), where M refers to the set of molecules $\{m_1 \ldots, m_n\}$, which is possibly of infinite size, R to the a set of n-ary operations/reaction rules $\{r_1 \ldots, r_n\}$ on the molecules, and A, which denotes an algorithm describing how to apply the rules R to a subset $P \subset M$. Each reaction rule $r_i \in R$ is written as a chemical reaction like

$$(x_1, x_2) \rightarrow^{r_i} x_1^* x_2^*$$

With that, we firstly introduce the molecular model, present several reaction rules to explain its depth, and demonstrate its strength on an example.

2 A SET OF MOLECULES

We understand each attribute D_i inside a database table D as a nucleus α_i that owns a *cloud of values* $e_1, \ldots e_k$ at distance ε_i. Each e_i corresponds to a data

Schommer C. (2010).
A MOLECULAR CONCEPT OF MANAGING DATA.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 411-415
DOI: 10.5220/0002758304110415

value $d_i \in D_i$ that might be e.g. of type *string* or *number* (integer, real, dots), but not a list of values (first normal form is valid). A nucleus α_i owns a *name* (= the attribute name) and shares a higher *valency* ν_{D_i}, the more dense the *cloud of values* is. The distance between the *nucleus* α_i and each *valectron* e_i gives the *strength of existence*, meaning that if the occurrence of e_i increases the occurrence of e_j, the distance to the nucleus is shorter. If the nucleus owns only one e_i, then $\alpha_i = e_i$.

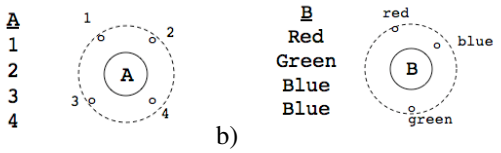

Figure 1: Database attributes D_A and D_B with the corresponding nuclei α_A and α_B.

Figure 1 presents two database attributes D_A and D_B with the corresponding nuclei α_A and α_B. Whereas all *valectrons* of nucleus α_A shares the same distance, the distance of e_{blue} of nucleus α_B is shorter than for e_{green} and e_{red}.

In opposite to its atomic basis, a database table of ≥ 2 attributes $D_1, \ldots, D_k$ is consequently a set of nuclei $\alpha_{D_1}, \ldots, \alpha_{D_k}$. The nuclei, however, are not organised in an arbitrary way, but keep themselves ordered:

- The lower the *valency* ν_{D_i} is the more centric the nucleus α_i will be.
- In case that some nuclei share the same *valency*, we may randomly select one of them.

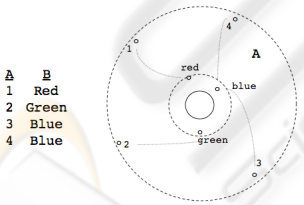

Figure 2: Simulation of two database attributes D_A and D_B of the database table D with the corresponding three-dimensional molecule $m_{A,B}$ (ordered by their *valency*).

The principle of ordering is unlike the ordering of a set of numbers but more the arrangement of the nucleus including their *cloud of values*. With respect to this, a chemical structure of size ≥ 2 – which is said as to be a *molecule* m∈M – therefore can not be a two-dimensional model anymore: the *cloud of values* embraces each previously selected nucleus and associates each *valectron* e_i with its corresponding partner of the other nucleus. Figure 2 shows a simulation of two database attributes D_A and D_B of the database table D with the corresponding three-dimensional molecule $m_{A,B}$ (ordered by their *valency*). As presented in Figure 2, the merge of nuclei is as follows:

- Assume that $\nu_{D_i} < \nu_{D_{i+1}} < \ldots \nu_{D_k}$, then ν_{D_i} has the highest priority and therefore takes over the innermost position, followed by $\nu_{D_{i+1}}$, and so on.
- The nuclei $\alpha_i, \ldots \alpha_k$ are nested and represented by their *cloud of values* only.
- Originally associated tuples inside D = $\{D_i, \ldots D_k\}$ are connected by molecular bridges $\gamma_{i,k}$ of a certain strength, which may vary.

3 ENZYMATIC REACTIONS

An artificial enzyme is a protein that is able to execute reactions. Whereas in the natural example an *enzyme* takes over the responsibility of many functions that concern the metabolism of an individual, the simulation of *enzyme* in a database environment can be understood as the adequate to reaction rules. Enzymatic reactions work in one or two ways

- the targeted nuclei α_i can be copied.
- the molecular bridges $\gamma_{i,k}$ between the *valectrons* may be destroyed.

but the enzymatic instruction decides if both or only the latter action takes place. For example, an enzyme that simply has to read existing molecules surely copies the existent structure and then keeps only those connections that satisfy the enzymatic instruction. On the other side, a permanent delete of data in the original molecule does not afford a copy but only the delete of the molecular bridges.

With respect to a retrieval, fundamental *enyzmes* concern the selection and projection enzyme. Given a molecule as presented in Figure 2, then the reaction rule $\sigma_{A,B}$ characterizes a chemical reaction of the original molecule – which consists of the two *cloud of values* A and B – to another molecule A^* B^*. The density and the valency change, since for example $\nu_{A,B} > \nu_{A^*B^*}$ of the new molecule:

$$A, B \rightarrow^{\sigma} A^* B^*$$

Similarly, the reaction rule π characterizes a chemical reaction as well, but in contrast to the reaction σ, the valency remains stable, whereas the number of resulting nuclei α_i changes:

$$A, B \rightarrow^{\pi} A^* B^*$$

As an example, Figure 3 shows an enzymatic reaction $\pi_B(\sigma_{B='blue'}(D))$, where the original molecule is copied and all molecular bridges and *valectrons* (except 'blue') are removed. Please note that the distance of ε_{blue} remains unchanged, i.e., the *valectron* remains at the same position.

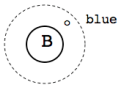

Figure 3: Examples of the Reactor Rule σ: select B from D where B='Blue'

Similarly, Figure 4 shows an enzymatic reaction $\pi_A(\sigma_{A>2}(D))$, where two *valectrons* occur.

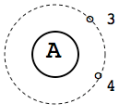

Figure 4: Examples of the Reactor Rule σ: select A fom D where A > 2

The enzymatic reaction $\pi_A(\sigma_{A>4}(D))$ given in Figure 5, however, gives only the nucleus α_A, but no *valetrons*.

Figure 5: Examples of the Reactor Rule σ: select A from D where A > 4

Finally, Figure 6 shows an enzymatic reaction $\pi_{A,B}(\sigma_{A=2}(D))$ where both nuclei α_A and α_B occur and the molecular bridge $\gamma_{'green',2}$ between the corresponding valectrons still exist.

With an enzymatic reaction $\in_{m,M}$, the insert of a new molecule *m* into an existing molecular structure M takes place by a simple addition. In case that *valectrons* are already present, these become merged. As an example, Figure 7 shows the merge of two molecules where the *valectron* e_3 is common. However, such an insert may violate the correctness of the existing data landscape, because it allows the creation of molecules that do not exist. The insert of the molecule

$$\in_{(e_5,e_6,e_3,e_4,e_9),M}$$

seems to be safe, but the existence of another molecule (e_1, e_2, e_3, e_7, e_8) causes an error, since inherently the molecules (e_5, e_6, e_3, e_7, e_8) and (e_1, e_2, e_3, e_4, e_9) might untruly be present as well. By using just one molecular bridge $\gamma_{i,j}$, we therefore risk the inconsistency of the whole molecule.

An insert, and moreover the presence of collections of *valectrons* must not have a single molecular bridge but a double one. With this, the dashed bridge

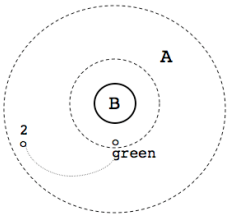

Figure 6: Examples of the Reactor Rule σ: select B from D where A = 2

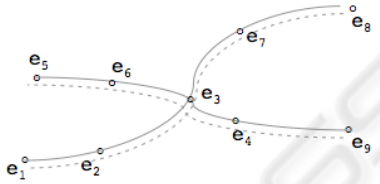

Figure 7: The reaction rule $\in(m,M)$: a molecular bridge (dashed) characterizes the connections of the *valectrons* (= the β-helix $\beta_{e_i...e_{e_j}}$) whereas the solid line the situation of the molecule after insertion (γ-helix).

characterizes the connections of the *valectrons*. This called the β-helix $\beta_{e_i...e_{e_j}}$. The solid line the situation of the molecule while insertion, representing values between the associated nuclei α_i. The molecular string is therefore called the α-helix. And with that, a molecule (e_5, e_6, e_3, e_7, e_8) does not exist since no α-helix is from $e_5 - e_6 - e_3$ to e_7.

As a third operation, an (equi-)join operation of molecules may be represented by the reaction rule $\lhd_=(M_i,M_j)$. As for the insert reaction $\in$, those valectrons, which occur both in molecule M_i and in M_j, are merged. All original valectrons keep their helix structure (see Figure 8).

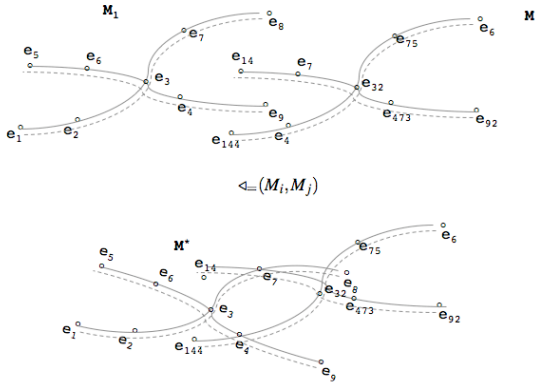

Figure 8: Reaction Rule $\lhd_=(M_i,M_j)$

With an enzymatic reaction $\notin_{m,M}$, we denote the delete of a molecule m within M. The helix $\beta_{e_i...e_{e_j}}$ guarantees that only those valectrons, which belong together, are deleted.

In addition, the composition of several reaction rules like

$$A, B \rightarrow^{\sigma,\pi} A^* B^*$$

is possible and appears in that order the reaction rules are given. The composition is commutative.

Beside the given reaction rule, the *authorization* of valectrons might be interesting as well. With *authorization*, we identify the an enzyme's right to access to a nuclei and it's cloud of values. This is not really a reaction rule as the enzymatic reaction does not results in a chemical reaction; it is more a feature of the nuclei itself that allows or disallows a permitted access. We therefore note a disallowed access by

$$\neg\alpha_A$$

meaning that the nucleus α_A rejects any kind of reaction. Instead of delivering a valectron, the result could be an empty element.

4 DISCUSSION

The idea of understanding data within an artificial chemical system is potentially unlike the relational system but offers a variety of characteristics. First, no data type specification is needed. The presence of a data item within the chemical database model is per se self-explaining and does not need any further specification concerning its type. The consequence then is that data (of different data type – from a relational point of view) is being identical. This is not of disadvantage because the expression of strength between valectrons through the molecular bridges $\gamma_{i,j}$ is very present. In fact, this is the second point as strong relationships among valectrons do inherently exist. If a combination of valectrons $e_i - e_j$ occurs often enough, then its molecular bridge $\gamma_{i,j}$ becomes stronger as if it occurs only "a few times". Third, the consideration of the molecular model towards a molecular-associative construct offers the identification of molecular clumps that are connected with each other and that represent a symbol, such that they may form a higher-related (cognitive) construct like a *mental image* or simply a *thought*. Assuming, that "tree" (for nucleus α_1), "green" (for nucleus α_2), and "rain" (for nucleus α_3) exist, it would certainly be possible to think of a "staying in the forest on a cold and rainy day". As a last point, the molecular data management model as described above is open for the input of data streams. Whereas the relational model lacks from high administrative efforts, a stream of data may be handled more effective in the proposed model.

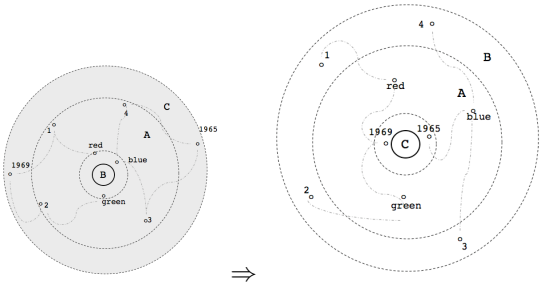

Figure 9: Restructuring the molecule: the left molecule refers to the situation where the number of years (α_C) is significantly less than the colour (α_A) and the amount of (α_B), whereas the right molecule refers to the more stable molecule.

On the other side, some kind of efforts is to be done in keeping the molecules in a stable and consistent form. *Stability* refers to a general claim that such nuclei α_i with a minor *valency* ν_i do more contribute to a general *model consistency* and therefore to the *stability* as well as those nuclei with a more dense*cloud of values*. In consequence of a delete or an insert of molecules, a restructuring reaction must take place in order to guarantee stability and consistency. With respect to this, assume that an insert of a new data leads to a change of the valency with $\nu_{D_i} > \nu_{D_{i+1}} < \ldots \nu_{D_k}$. Then, the enzymatic restructuring ψ is as follows:

- A copy α_i' of the nuclei α_i is created; it is then set on its new place, depending on its valency ν_{D_i}.
- All valectrons e_i of α_i walk on the β-helix β_i and finally reach their *cloud of values*.
- At each point, a connection of each valectron remains.

On the other side, a continuous change of the number of values may become counter-productive and finally refer to a continuous and repeating restructuring of the molecule, such that nuclei are more concerned with internal configurations than with the management of data. An alternative therefore is to prefer those nuclei whose *cloud of values* do not or even less changes in size. Once the molecule is created (first approach) and once a certain information about stable nuclei have been got, the second solution seems to be more appropriate.

5 CONCLUSIONS

With the presented proposition, we follow the concept of understanding data and information as an (artificial) chemical model. Each data is consistent with

an atomic entity but each combination of data corresponds to a molecular structure. An attribute D inside a relational system is represented as a nucleus α_D sharing a *cloud of values*, which consists of so-called *valectrons*. The nucleus satisfies the first normal form (atomic values). By using reaction rules like the selection of tuples σ or projection of attributes π, a retrieve of molecules can be achieved quite easily. As mentioned in chapter 4, one of the major advantages is the associative nature of the molecule. The generation of mental images (or thoughts), beside the implementation of the given system, will then be next steps.

ACKNOWLEDGEMENTS

This work is currently been done within the International Laboratory for Intelligent and Adaptive Systems of the University of Luxembourg. We thank the members of the MINE research group for their support.

REFERENCES

Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial chemistries-a review. *Artificial Life*, 7(3):225–275.

Fernández-Baizán, M. C., García, A., González, M. M., Pérez-Llera, C., Portaencasa, R., and Santos, E. (1996). Analysis and design of a relational database management system and implementation of its nucleus. *Computers and Artificial Intelligence*, 15(4).

Gerrilsan, R. (1975). The application of artificial intelligence of data base management. In *IJCAI*, pages 521–527.

Hutton, T. J. (2002). Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life*, 8(4):341–356.

Kelemen, J. and Sosík, P., editors (2001). *Advances in Artificial Life, 6th European Conference, ECAL 2001, Prague, Czech Republic, September 10-14, 2001, Proceedings*, volume 2159 of *Lecture Notes in Computer Science*. Springer.

Leach, A. (2001). *Molecular Modelling - Principles and Applications*. Prentice Hall, 2nd edition.

Schommer, C. (2009). An artificial molecular model to foster communities. In *Knowledge Discovery and Information Retrieval (KDIR)*. IEEE Computer Society.

Skusa, A., Banzhaf, W., Busch, J., Dittrich, P., and Ziegler, J. (2000). Künstliche chemie. *KI*, 14(1):12–19.

Tominaga, K., Watanabe, T., Kobayashi, K., Nakamura, M., Kishi, K., and Kazuno, M. (2007). Modeling molecular computing systems by an artificial chemistry - its expressive power and application. *Artificial Life*, 13(3):223–247.

von Luck, K. and Marburger, H., editors (1994). *Management and Processing of Complex Data Structures, Third Workshop on Information Systems and Artificial Intelligence, Hamburg, Germany, February 28 - March 2, 1994, Proceedings*, volume 777 of *Lecture Notes in Computer Science*. Springer.

Ziegler, J. and Banzhaf, W. (2001). Evolving control metabolisms for a robot. *Artificial Life*, 7(2):171–190.