Molecules of Knowledge: a Novel Perspective over Knowledge Management

Stefano Mariani
Supervisor: Andrea Omicini

Alma Mater Studiorum – Università di Bologna
via Venezia 52 – 47521 Cesena, Italy
s.mariani@unibo.it

Abstract. To face the challenges of knowledge-intensive environments, we investigate a novel self-organising knowledge-oriented (SOKO) model, called *Molecules of Knowledge* (MoK for short). In MoK, knowledge atoms are generated by knowledge sources in shared spaces – compartments –, self-aggregate to shape knowledge molecules, and autonomously move toward knowledge consumers.

1 The *Molecules of Knowledge* Model

The *Molecules of Knowledge* model is a biochemically-inspired coordination model exploiting and promoting self-organisation of knowledge. The basic motivation behind MoK is the idea that knowledge should autonomously aggregate and diffuse to reach knowledge consumers rather than "be searched". The main pillars of MoK are represented by the following (bio)chemical abstractions:

atoms — the smallest unit of knowledge in MoK, an *atom* contains information from a *source*, and belongs to a *compartment* where it "floats"—and where it is subject to its "laws of nature";
molecules — the MoK units for knowledge aggregation, *molecules* bond together somehow-related atoms;
enzymes — emitted by MoK *catalysts*, *enzymes* influence MoK reactions, thus affecting the dynamics of knowledge evolution within MoK compartments to better match the catalyst's needs;
reactions — working at a given *rate*, *reactions* are the biochemical laws regulating the evolution of each MoK compartment, by ruling knowledge aggregation, diffusion, and decay within compartments.

Other relevant MoK abstractions are instead in charge of important aspects like topology, knowledge production and consumption: **compartments** represent the conceptual *loci* for all MoK entities as well as for biochemical processes – like knowledge aggregation and diffusion –, providing MoK with the notions of *locality* and *neighbourhood*; **sources** are the origins of knowledge, which is continuously injected in the form of MoK atoms at a certain *rate* within the compartment sources belong to; **catalysts** stand for knowledge *prosumers*, emitting enzymes which represent their actions which affect knowledge dynamics

within their own compartment, especially to increase the probability of providing him/her with relevant knowledge items.

Atoms Atoms are the most primitive living pieces of knowledge within the model. A MoK atom is produced by a knowledge source, and conveys a piece of information spawning from the source itself. Hence, along with the content they store, atoms should also store some contextual information to refer to the content's origin, and to preserve its original meaning.

As a result, a MoK atom is a triple of the form $atom(\ src,\ val,\ attr\ )_c$ where: *src* identifies unambiguously the source of knowledge; *val* is the actual piece of knowledge carried by the atom – any kind of content –; *attr* is essentially the content's attribute, that is, the additional information that helps understanding it – possibly expressed according to some well-defined ontology or controlled vocabulary –; c is the current *concentration* of the atom, which is essentially the number of atoms of the kind in the compartment.

Molecules Molecules are spontaneous, stochastic, "environment-driven" aggregations of atoms, which in principle are meant to reify some semantic relationship between atoms, thus possibly adding new knowledge to the system—for instance self-aggregated news chunks could shape the conceptual "path" toward a novel interesting news story. Each molecule is simply interpreted as a set of atoms, that is, an unordered collection of somehow semantically-related atoms.

A MoK molecule has then a structure of the form $molecule(\ Atoms\ )_c$ where c is the current concentration of the molecule, and *Atoms* is the collection of all the atoms currently bonded together in the molecule.

Enzymes One of the key features of MoK is that it interprets prosumer's knowledge-related actions as *positive feedbacks* that increase the concentration of related atoms and molecules within the prosumer's compartment, producing the positive feedback required to enable self-organisation of knowledge.

A MoK enzyme has then a structure of the form $enzyme(\ Atoms\ )_c$ where every enzyme – with its own concentration c – explicitly refers to a collection of *Atoms* that the catalyst's actions have in any way pointed out as of interest for the catalyst him/herself.

Biochemical Reactions The behaviour of a MoK system is actually determined by the last abstraction of the model: the *biochemical reaction* [1].

As a knowledge-oriented model, the main issue of MoK is determining the semantic correlation between MoK atoms. So, to define a working MoK system, the basic $mok(\ atom_1,\ atom_2\ )$ function should be defined, which takes two atoms $atom_1$, $atom_2$, and returns a *"matching degree"*—which could be a boolean or a double value $\in [0, 1]$. The precise definition of *mok* depends on the specific application domain.

The *aggregation reaction* (AggR) bonds together atoms and molecules:

$$molecule(Atoms_1)\ +\ molecule(Atoms_2) \longmapsto^{r\text{-}agg} molecule(Atoms_1 \bigcup Atoms_2)\ +\ Residual(Atoms_1, Atoms_2)$$

where r_agg is the AggR reaction rate, $mok(atom_1, atom_2)$ holds for some $atom_1 \in Atoms_1$, $atom_2 \in Atoms_2$, and $Residual(Atoms_1, Atoms_2)$ is the multiset of atoms obtained as the multiset difference $(Atoms_1 \uplus Atoms_2) \backslash (Atoms_1 \bigcup Atoms_2)$—that is, essentially, all the atoms of $Atoms_1$ and $Atoms_2$ that do not belong to the resulting molecule. In short, more complex molecules are formed by aggregation whenever some atoms in the reacting molecules (or in reacting atoms, handled here as one atom molecules) are semantically correlated (via the *mok* function).

Positive feedback is obtained by the *reinforcement reaction* (ReinfR), which consumes a single unit of enzyme to produce a unit of the relative atom/molecule:

$$enzyme(Atoms_1) \; + \; molecule(Atoms_2)_c \longmapsto^{r\_reinf} molecule(Atoms_2)_{c+1}$$

where r_reinf is the ReinfR reaction rate, $enzyme(Atoms_1)$ and $molecule(Atoms_2)_c$ exist in the compartment, both with $c \neq 0$, and $mok(atom_1, atom_2)$ holds for some $atom_1 \in Atoms_1$, $atom_2 \in Atoms_2$

Following the biochemical metaphor, molecules should fade as time passes, lowering their own concentration according to some well-defined *decay-law*. The temporal *decay reaction* (DecayR) is hence defined as follows:

$$molecule(Atoms)_c \longmapsto^{r\_decay} molecule(Atoms)_{c-1}$$

Similarly, a biochemical-inspired knowledge management model should provide some spatial interaction pattern. MoK adopts *diffusion* as its data-migration mechanism, in which atoms and molecules can move only between *neighbour* compartments, resembling membrane crossing among cells. MoK *diffusion reaction* (DiffR) is then modelled as follows, assuming that σ identifies a biochemical compartment, and $\|\|_\sigma$ brackets molecules in a compartment σ:

$$\begin{array}{c} \| \; Molecules_1 \bigcup molecule_1 \; \|_{\sigma^i} \; + \| \; Molecules_2 \; \|_{\sigma^{ii}} \longmapsto^{r\_diffusion} \\ \| \; Molecules_1 \; \|_{\sigma^i} \; + \| \; Molecules_2 \bigcup molecule_1 \; \|_{\sigma^{ii}} \end{array}$$

where σ^i and σ^{ii} are neighbour compartments, $r\_diffusion$ is the diffusion rate, and $molecule_1$ moves from σ^i to σ^{ii} as the result of the reaction.

2 News Management: A First Case Study

While MoK is a general-purpose model for knowledge self-organisation, it can be tailored on specialised application scenarios, by refining the notions of atom and suitably defining the *mok* semantic correlation function. Since news management provide a prominent example of a knowledge-intensive environment, we chose news management as the first case study for MoK, introducing the MoK-News model for self-organisation of news.

2.1 Knowledge representation for news management

IPTC[1] develops and maintains technical standards for improved news management, such as NewsML[2] and NITF[3].

[1] `http://www.iptc.org/`
[2] `http://www.iptc.org/site/News_Exchange_Formats/NewsML-G2/`
[3] `http://www.iptc.org/site/News_Exchange_Formats/NITF/`

The NewsML tagging language is a media-type orthogonal news sharing format standard aimed at conveying not only the core news *content*, but also the data that describe the content in an abstract way, namely the *metadata*. In order to ease syntactical and semantical interoperability, NewsML adopts XML as the first implementation language for its standards and for maintaining sets of *Controlled Vocabularies* (CVs), collectively branded as *NewsCodes*, to represent concepts describing and categorising news objects in a consistent manner—pretty much as domain-specific ontologies do.

The News Industry Text Format, too, adopts XML to enrich the content of news articles, supporting the identification and description of a number of news typical features, among which the most notable are: *Who* owns the copyright to the item, who may republish it, and who it's about; *What* subjects, organisations, and events it covers; *When* it happened, was reported, issued, and revised; *Where* it was written, where the action took place, and where it may be released; *Why* it is newsworthy, based on the editor's analysis of the metadata. NewsML in fact provides no support for any form of inline tagging to add information to the plain text, for instance with the purpose to ease the work of a text mining algorithm usable to automatically process the document. Thus, NITF and NewsML are complementary standards, hence they perfectly combine to shape quite a comprehensive and coherent framework to manage the whole news lifecycle: comprehensive, given that one cares about news overall structure, including metadata, whereas the other focusses on their internal meaning making it unambiguous; coherent, because they both exploit the same IPTC abstractions—in fact NITF, too, uses NewsCodes.

2.2 Towards MoK-News

Since sources provide journalist with the required *raw information* already formatted according to the afore-mentioned IPTC standards, a simple-yet-effective mapping can be drawn. In fact, MoK atoms do actually have a clear counterpart in NewsML and NITF standards: *tag*. Tags – along with their "content" – can in fact be seen as the atoms that altogether compose the "news-substance" in the news management scenario. As a result, our MoK-coordinated news management system would contain $< newsItem >$ atoms, $< person >$ atoms, $< subject >$ atoms, etc.—that is, virtually one kind of atom for each NewsML/NITF tag.

A MoK-News atom looks like $atom(\ src,\ val,\ sem(tag, catalog)\ )_c$, where

src ::= *news source uri*
val ::= *news content*
$attr$ ::= $sem(tag,\ catalog)$
  tag ::= *NewsML tag* | *NITF tag*
  $catalog$ ::= *NewsCode uri* | *ontology uri*

Here, the content of an atom is mostly given by the pair $< val, tag >$, where tag could be either a metadata tag drawn from NewsML or an inline description tag taken from NITF. The precise and unambiguous semantics of the new content (val) can be specified thanks to the $catalog$ information, which could be grounded in either NewsML or NITF standards in the form of NewsCodes, or

instead be referred to a custom ontology defined by the news worker. MoK-News molecules, enzymes, and biochemical reactions are then both syntactically and semantically affected by such domain-specific mapping of the MoK model.

3 Related & Future Works

To the best of our knowledge, althought the (bio-)chemical metaphor is widely used to achieve self-organisation and self-adaptation by emergence, no MoK-like approaches were studied that could bring self-* behaviours directly *into data*. On the other hand, in the news community much attention is paid to interoperability and semantic standardisation—but no *paradigm shift* has been tried to "see" data as active entities.

Nevertheless, some MoK-similar system actually exists although with different aims. In [2] a general-purpose, tuple-space-based approach to knowledge self-organisation was built exploiting a WordNet ontology to identify relationships between knowledge chunks – tuples –, and drive their migration to build clusters. In [3] a similar clustering behavior is achieved with *collective sort*, assuming that a $1:1$ relation exists between admissible tuple templates and tuple "sorts"—essentially making the number of clusters known *a priori*. The MoK approach is different in that it pushes the above cited "similarity-based clustering" to the limit: MoK not only aggregates somehow-related knowledge chunks in a same spot – e.g. diffusing news to interested journalists – but also tries to *physically merge* units of information to create new knowledge—the molecules.

Our future efforts will be on first devoted to provide an effective implementation of the MoK model upon an existing coordination middleware enriched with the "online biochemical simulator" behavior [4], then to test the implementation on the MoK-News application scenarios, and on other knowledge-intensive environments as well—e.g., *MoK-Research* and *MoK-HealthCare*.

References

1. Viroli, M., Casadei, M., Nardini, E., Omicini, A.: Towards a chemical-inspired infrastructure for self-* pervasive applications. In Weyns, D., Malek, S., de Lemos, R., Andersson, J., eds.: Self-Organizing Architectures. Volume 6090 of LNCS. Springer (July 2010) 152–176
2. Pianini, D., Virruso, S., Menezes, R., Omicini, A., Viroli, M.: Self organization in coordination systems using a WordNet-based ontology. In Gupta, I., Hassas, S., Jerome, R., eds.: 4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO 2010), Budapest, Hungary, IEEE CS (27 September–1 October 2010) 114–123
3. Gardelli, L., Viroli, M., Casadei, M., Omicini, A.: Designing self-organising MAS environments: The collective sort case. In Weyns, D., Parunak, H.V.D., Michel, F., eds.: Environments for MultiAgent Systems III. Volume 4389 of LNAI. Springer (May 2007) 254–271
4. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry **81**(25) (1977) 2340–2361