# Principles of Digital Evolution Machines

Istvan Elek
*Eotvos Lorand University, Budapest*
*elek@map.elte.hu*

## Abstract

*This paper introduces a theoretical approach of the construction of a self developing and adaptive artificial digital organism with huge remembrance and the ability of the interpretation of the surrounding world. The paper describes the self development of intelligence of digital organisms from small fragments of digital knowledge.*

## 1. Introduction

There is no right definition of intelligence. Instead of making useless definitions, it is more promising to set up some essential principles that adjust the process of collecting and interpreting data from the surrounding world. There are many research papers that describe machines which collect data from the surroundings and they have some kind of remembrance [5]. Their ability of interpretation of the environment is restricted. The limits of these constructions are obvious: their intelligence never becomes similar to that of mammals or octopuses.

Many papers [6], [7], [1] dealing with artificial organisms emphasized the importance of complexity. Not only life shows serious complexity. There are human made constructions as well, such as the topology of Internet, that have increasing complex structures with fractal properties [3], [2], [4].

This paper establishes some principles that are really simple but like bricks they can be combined into arbitrarily complex buildings. The purpose is to construct a self developing and adaptive artificial digital organism, a so called digital evolution machine (**DEM**) that collects and arranges, sometimes even restructures its database.

### Some aspects of the evolution of life

Paleontology and geology serve many exciting examples of the one way evolution. The time flows in one direction, forward. Life always tries to adapt itself to the circumstances, mainly to the weather (temperature, seasons, climate). If the climate has changed, the adaptive organisms also change their right properties, skills and manners. If one million years later the climate changed back, the evolution did not remember the previous stage of the environment. The adaptivity also produces new properties that help to survive the changes. The evolution seems to be recursive. It means the regulation is a kind of recursion, like a feedback. The current stage is the basis of the next step. Evolution never gets back. There are no general physical laws in the background where the processes can be computed from.

If an organism did not recognize the enemy, the food and the environment, it had died soon or immediately. Consequently, organisms had to collect data from their surroundings (perception) and interpret them (understanding). The remembrance is probably one of the most important components of understanding the world. The unreversable time, i.e. the serial recording of events and data of the environment produces a huge database. It has to contain everything that happened to an organism in its lifetime. The key of surviving is the interpretation of the surrounding world and the action. The speed of the reaction is also important among animals.

### An ancien story from the early era of evolution

Look at the following event that took place in the ancient ocean many million years ago. A very primitive organism was swimming in the water. The temperature of the sea became somewhat colder for him. He detected the low temperature and realized it was bad for him. He decided to swim toward warmer waters (Figure 1).

The question is whether this reaction involves intelligence or not. A quick look says, yes, it does. A thermostat does the same: detects, compares, makes a dicision and acts something, but it is not yet considered intelligent, because a kind of remembrance requires a huge knowledge-base and a fast graph algorithm.

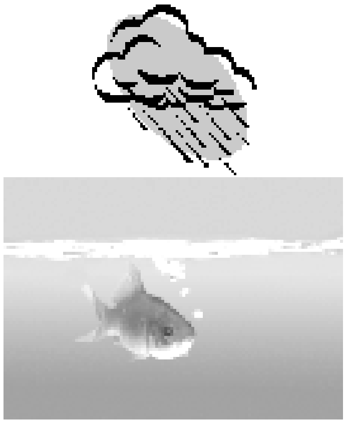

Figure 1. The weather and climate were the most effective factors of the evolution of organisms. Every organism needs to interpret the measured and stored data of the environment if they wanted to survive. This ability required a huge database containing everything that ever happened to the organism. This is experience. The interpretation required a fast graph algorithm that could find the right answer to the current challenge.

### The construction of the knowledge base

Biologists say there is a kind of self organisation among protein molecules in vitro. Even if we assume it was the first step toward a real organism, it is evident there is no huge knowledge database and complicated interpretation logic in it. Protein molecules seem to be inclined to form combinations. Their data collection logic also has to be simple. Consequently a knowledge database can be constructed from simple steps of data collection. The algorithm of the knowledge graph search has to be simple.

Look at a traditional knowledge database for nature sciences.This construction is hierarchical like a tree constructed by scientists (Figure 2).

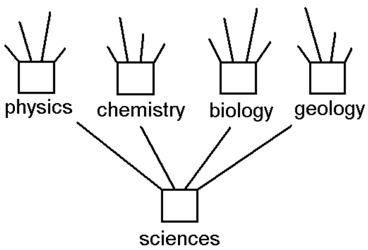


Figure 2. The hierarchy of the natural sciences is a kind of knowledge-base which forms a tree.

How to construct a knowledge base for a simple organism that has no scientists? The following sections try to answer this question.

## 2. Principles of an organic knowledge-base

Let us see some essential principles that regulate the data collection (perception) for an organism that is named DEMentity in the followings.

### General principles

1. Data collection process is a lifelong task that builds up the knowledge base.
2. Let knowledge elements be named atoms.
3. Atoms are situated in a tree structure that is a graph.
4. There is no lonely atom, every atom connects to another one at least.
5. Any atom can be a member of any group. A group of complex knowledge elements is named a context.
6. Identical operations are valid for atoms and contexts.
7. Let us allow contradictory atoms in the knowledge base.
8. Let us name a path the trajectory between two or more atoms.
9. If a search of the graph produces succes, the path becomes stronger (imprinting). If a search produces fail, the path becomes weaker (oblivion).

### Quantification of the knowledge and atomic distances

Let us construct the knowledge-graph (Figure 3). It consists of atoms and their connections. Atoms are the nodes and connections are the edges of knowledge-graph. Let us define a context that includes arbitrary atoms of the tree; consequently, the structure of the graph is not predefined. It depends on the atomic connections that depend on a time series, when events took place in time order one after the other.

The general principles declared the equality of atoms and contexts. In other worlds a context can contain simple atoms or other contexts as well.

Let $a_i^j$ denote the $i$th atom in the $j$th context which contains $N$ atoms. Let the quantity of the knowledge of a context ($k^j$) be the sum of the quantity of the knowledge of every atom in it:

$$k^j = \sum_{i=1}^{N} a_i^j \qquad (1)$$

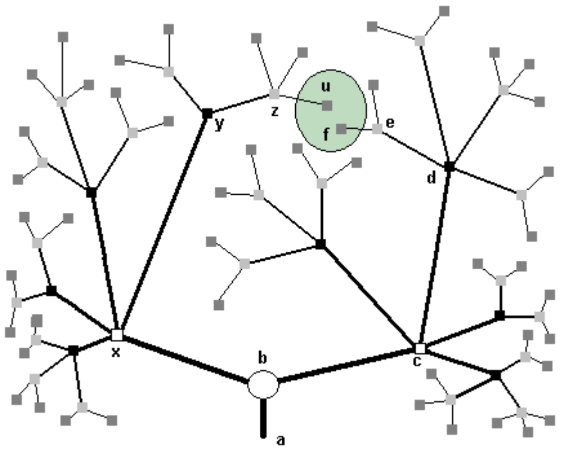


Figure 3. The Knowledge-graph

The knowledge-base has two basic functions: receiving a question and answering. The key of the problem is to find the path from the questioning node to the answering node in the knowledge-base.

Let $l_{ij}$ denote the strength of the connection between $a_i$ and $a_j$ . Let $l_{ij} = l_{ij} +1$ if the tour produces good result and $l_{ij} = l_{ij} -1$ if the result is bad. This logic makes good paths stronger and bad ones weaker.

Look at the *u* and *f* nodes in the knowledge-graph (Figure 3). What is the right distance definition for them? In the case of Euclidean distance, *u* and *f* nodes are near. Since there is no direct connection between them, a better distance definition has to depend on the length of the path along branches.

Let $a_i$ and $a_j$ be two nodes of the knowledge-graph where the path includes $m = j - i$ atoms between them. Let $d_{ij}$ be the distance of these two nodes, let the strength of their connection be denoted by $l_{ij}$ , which is the reciprocal of the sum of the strength of the connections between $a_i$ and $a_j$ . The stronger the connection between two nodes, the closer they are.

$$d_{ij} = 1/\left(\frac{\sum_{k=i}^{j-1} l_{k(k+1)}}{m}\right) \quad (2)$$

The goal of the knowledge-graph is to answer a question arrising from the circumstance. How to find the rigth path from the questioning node to the answering one? The fastest is the right path, probably. This logic produces very fast reaction in well known problems and may result fail in unknown cases. The fail means unsuccessful escape, capture or something important for the organism. If it survived the situation, i.e. the reaction was successful, and the path that produced the success became stronger. If it did not survive the situation or the result of the action was failed, i.e. the result of the action was unsuccessful, the organism was knocked out or a path in the knowledge-graph became weaker.

## Non hierarchical knowledge-base

Let **K** denote the knowledge-base which consists of $n$ atoms $a_i$, $a_j \in$ **K**. Let us name it the knowledge-matrix.

$$\mathbf{K} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad (3)$$

Some of the atoms are in touch with other atoms in **K**. Let us describe the links of $a_i$ and $a_j$ atoms with $l_{ij}$ , where

$$l_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \text{ and no link between them} \\ u - v & \text{else where } u \text{ successful and } v \text{ unsuccessful} \end{cases}$$

Let us organize the atomic links into a matrix form, and name it a link matrix and denote it by **L**. The elements of the link matrix are $l_{ij}$ that describe the link of the atomic pointpairs. **L** is diagonal ($l_{ii} = 1$) and describes the relationships of the atoms in the knowledge-matrix:

$$\mathbf{L} = \begin{pmatrix} 1 & l_{12} & \dots & l_{1n} \\ l_{21} & 1 & \dots & l_{2n} \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix} \quad (4)$$

## The properties of the K and L matrices

1. Every atom has one link at least.
2. Let **C** be a context. **C** $\subseteq$ **K**, i.e. **C** consits of any atoms of **K**.
3. Any atoms can be the member of any context.
4. Any contexts can be a member of any contexts. In this case, the knowledge-matrix (**K**) is a hyper matrix where matrix elements can be matrices.
5. In summary an atom can be a
    a) Simple atom that is really elementary and belongs to one context.
    b) Multi-member atom that is also elementary but belongs to more than one context.
    c) Aggregated atom that is a kind of simple atom, but its value is a representation of a context. In other words, its value is a determinant or a spur of the knowledge-matrix of a certain context.

d) A complex atom that is a context that includes any kinds of atoms above (Figure 4).

### The evolutional snippets of the knowledge-base

1. The knowledge-base is a continuously increasing database which stores everything that happened to it. This is a one way process. The knowledge-base is different from entity to entity.
2. There is no changing or erasing function in the knowledge-base.

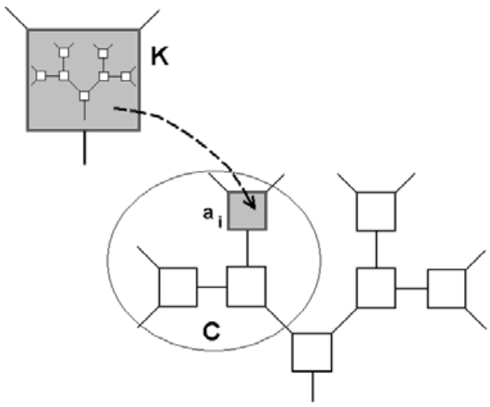


Figure 4. A context is a complex atom in the knowledge-base. Any arbitrarily complex knowledge-base can be constructed based on this logic. Since the knowledge-base is a continuously increasing dataset, links can be arbitrarily complex too. In this case, the context K is a complex atom ($a_i$) of the context C.

3. Let the data collection be *extensive* if a DEM-entity perceives the circumstance and stores data. There are some important consequences of the extensive mode:
   a) Every individual knowledge-base is different. It depends on the career of a certain DEM-entity. There are as many DEM-entities as many kinds of knowledge-base exist.
   b) If some changes happen that certainly produces the same circumstance in the past, the previously recorded atoms has not been changed, simply a new atom has appended the end of the knowledge-base.
4. Let us have another mode that was named *intensive*, when there is no perception. It is a „meditative" stage when the knowledge-base acquires the data came from the extensive stage. The result of the data acquisition in intensive mode may produce new contexts, faster path, more reliable work. This stage is extremely important in learning the circumstance.
5. The feedback is a process when a DEM-entity is informed about the result of its reaction. The result of this process is success or fail. As mentioned previously, the success/fail makes stronger/weaker a certain path in the knowledge-graph of the knowledge-base. The DEM-entity's knowledge-base becomes much stronger/ weaker if the feedback such as rewarding or punishment comes from an external intelligent entity, because its knowledge-base can be considered as an included context.
6. In the history of the Earth there was never only one organism. There were always ensembles. Ensembles make organisms competitive. Competition results in different skills i.e. different knowledge-bases.
7. Different circumstances cause different experiences for the organisms.

### Searching in the knowledge-base

It is an elementary function to find the right answer quickly if a question appears. The knowledge-graph contains the data that are the source of answering. How to find the right answer?

1. Regarding the early era of the evolution where the combinations of protein molecules were the most complex structure in the world, we can not suppose the existence of fascinating graph algorithms in this structure. Only simple algorithm can be like the preference of the fastest path, the most stronger path, or something like that. It is named preconception in the everyday life. The preconception accelerates getting answer.
2. So there is no more complicated algorithm that answer a certain question. While the knowledge-base is increasing, the graph structure becomes more and more complicated. The same simple algorithms have to serve the right answer.
3. Questions and answers have to be stored together in the knowledge-base. Questions may identify the start up context where the answer can be found. This is also a kind of preconception.
4. The judgement of an answer of the organism (or DEMentity) comes from the environment. If the answer is right, the organism confirms the right path in the knowledge-graph, if not the path has weakened. The result is a more developed knowledge-base.

## 3. Some constructional aspects of the knowledge-base

The evolution is a process in time and in space dimensions. Regarding the results of paleontology it is known that evolution is recursive, i.e. an evolution step depends on the previous step only and it is a one way process, it can not be turned back. This is known as Dollo's law in paleontology [8]. This law was first stated by Dollo in this way: „An organism is unable to return, even partially, to a previous stage already realized in the ranks of its ancestors". According to this hypothesis, a structure or organ that has been lost or discarded through the process of evolution will not reappear in that line of organisms.

The development means not only physical but mental changes as well in an organism. Mental means intelligence in this context. The adaptivity comes from its knowledgebase. The quality of the knowledge-base has to influence the physical properties also. Consequently the physical and mental evolution work collateraly.

## Synthetic worlds, artificial environment

If we are willing to create experienced DEM-entities there is no wizard unfortunately. There is no a recipe to install them from little pieces. How to construct a DEM-entity? Probably, it is impossible to construct only one of it. First we should create an ensemble from many initial DEM-entites and leave them to live in their own world.

We have two access points to this problem. The first one is to construct an artificial circumstance where DEM-entities live in. The second one is to construct many initial DEMentities with simple perception and interpetation functions. Let us look at some of the details:

1. The artificial world (AW) has some essential properties that define the frame of this world.
   a) Let AW be huge where circumstances have spatial dependencies. Regarding the size of this world, environmental parameters are obviously different. If we leave DEM-entities alone in this world they will have different experiences and different knowledge-bases because of climatic differences.
   b) If the AW is huge enough, the survival strategies will be different. One of the DEM-entities escapes from the unfriendly circumstances but others try to adapt. Different strategies result different knowledge-bases.
2. If there are many DEM-entities on the same territories, what is the consequence?
   a) There are many DEM-entities who try to get better strategy in order to be more successful than others. Someone gets advantages but someone gets disadvantages since it fails to answer a certain question.
   b) Regarding the different DEM-entities and unique knowledge-bases, many different strategies can coexist in the AW. Consequently, many different strategies can be successful at the same time. Someone preferes the escape, but someone the competition.
3. Many DEM-entities will have many different knowledge-bases.

## The Digital Evolution Machine

The question is how to construct the prototype of a DEMentity? Before the construction of the prototype, let us create the artificial world that will be the space for DEM-entities. If AW has been created already many DEM-entities should be available to start up in it. Properties and abilities of DEM-entities were introduced previously, so the task is to make their software representation. Regarding the quantity of DEMs and the huge sized world with different spatial properties may result many formed DEMs, and these entities have different knowledge-bases.

An intelligent DEM-entity can exist in ensemble only. There is no lonely intelligency because it is a collective product that is realized in some successful entities' knowledgebases.

## Pending questions

There are many pending questions unfortunately. Let us see some of them:

1. The existence of the extensive stage is obvious, but not the intensive one. The existence of the intensive stage does not result from the principles mentioned above.
2. The next unresolved problem is related to the intensive stage as well. The instensive stage is a process when the perceived data coming from the extensive stage have been evaluated and restructured into existing or new contexts. Why and how do the evaluation produce a new context? Which principle made the knowledge-base restructured?
3. It would be very useful if there were a lifetime for a DEM-entity. DEMs collect data every day, every minute and the knowledge-base is building up. If the starting date of data collection is too far from now the knowledge-base has many unimportant and useless information even contexts. Let us see Figure 5. Is there any value for me if I know how to conquer a knight in tournament? No, it is absolutely indifferent to me because I don not have any sword, horse and there is no knights ever. Unfortunately the necessity of the limited lifetime does not come from the principles.

Figure 5. Don Quijote and his attendant Sancho Panza. Before Don Quijote's era knights' skills in tournament was essential for a knight, but today it is not more than a hobby. [Daumier's draw, 1860]

## 4. Conclusions

This paper tried to introduce some principles and suppositions of how to construct adaptive and self-developing digital evolution machines having their own knowledge-bases that help them to understand and survive the challenge of the environment. This is only a theoretical approach until we do not develop the huge artificial world and millions of DEMentities. The monitoring of DEMs' life will show the way of the evolution of intelligence.

## References

[1] C. Adami – C. Ofria – T. Collier: „Evolution of Biological Complexity", arXiv:physics/0005074v1 26 May 2000

[2] Albert, Barabasi: „Statistical mechanics of complex networks", Reviews of modern physics, VOLUME 74, January 2002

[3] Barabasi – Albert – Jeong: „Mean-field theory for scale-free random networks", Preprint submitted to Elsevier Preprint, 5 July 2002.

[4] M. Newman – A-L. Barabasi – D. J. Watts: „The structure and Dynamics of Networks", Princeton University Press, 2006

[5] S. Russel – P. Norvig: „Artificial Intelligence: A Modern Approach", Prentice Hall, 2002

[6] C. Adami: „Ab Initio of Ecosystems with Artificial Life", ar-Xiv:physics/0209081 v1 22 Sep 2002

[7] C. Adami: „Sequence Complexity in Darwinian Evolution", Complexity, 2003, Wiley Periodicals, Vol.8, No.2

[8] Louis Dollo: http:/en.wikipedia.org/wiki/Dollo's law