Content Aggregation, Visualization and Emergent Properties in Computer Simulations

Gordana Dodig Crnkovic¹, Juan M. Duran² and Davor Slutej³

¹School of Innovation, Design and Engineering, Computer Science Laboratory, Mälardalen University, Sweden. gordana.dodig-crnkovic@mdh.se
²SimTech, Universität Stuttgart, Deutschland and University of Córdoba, Argentina. juan.duran@philo.uni-stuttgart.de
³ABB and Mälardalen University, Sweden, davor.slutej@se.abb.com

Abstract
With the rapidly growing amounts of information, visualization is becoming increasingly important, as it allows users to easily explore and understand large amounts of information. However the field of information visualization currently lacks sufficient theoretical foundations. This article addresses foundational questions connecting information visualization with computing and philosophy studies. The idea of multiscale information granulation is described based on two fundamental concepts: information (structure) and computation (process). A new information processing paradigm of Granular Computing enables stepwise increase of granulation/aggregation of information on different levels of resolution, which makes possible dynamical viewing of data. Information produced by Google Earth is an illustration of visualization based on clustering (granulation) of information on a succession of layers. Depending on level, specific emergent properties become visible as a result of different ways of aggregation of data/information. As information visualization ultimately aims at amplifyng cognition, we discuss the process of simulation and emulation in relation to cognition, and in particular visual cognition.

Introduction
Chen remarks in a newly published article on information visualization:

“The general consensus, as reported by a recent workshop and a few other public presentations, was that information visualization currently lacks adequate theoretical foundations. [Nor07] As a result, many approaches are ad hoc in nature. […] The search for theoretical foundations increasingly introduces and adopts theories and conceptual frameworks from other fields and disciplines. For example, distributed cognition in human-computer interaction is seen as a potential candidate for a theoretical framework for information visualization. [Sta08] Norman’s Seven Stages of Action, also in human-computer interaction, provides a new insight into interacting with information visualizations, specifically on the gulf of execution and the gulf of evaluation. [Lam08]” http:\onlinelibrary.wiley.com/doi/10.1002/wics.89/full, [Che10].

This article offers a contribution to the theoretical foundations of information visualization from the point of view of computing and philosophy studies, providing broader context in which information visualization is related to cognition and emergent properties in physical and simulated worlds.

Info-computational View of Cognition
All of our knowledge is based on information we get from the world. Physicists [Zei05] [Llo06] and [Ved10] suggest possibility of seeing information and reality as one. This is in accord with Informational Structural Realism which says that reality is made of informational structures [Flo08a]. What Floridi assumes to be mind-independent data corresponds to the world as information (proto-information). By interactions with an agent it reveals as consisting of structural objects (chunks of information with defined mutual relationships).

Building on Floridi’s Informational Structural Realism with information as the fabric of the universe the process of dynamical changes of the universe makes the universe a huge information processing mechanism. It is important to understand that computation performed by the Universe is not the computation we have in our machines, it is com-
putation most generally defined as information processing, [Bur05]. Here is how Chaitin describes the idea of computing Universe (Natural computationalism):

“And how about the entire universe, can it be considered to be a computer? Yes, it certainly can, it is constantly computing its future state from its current state, it’s constantly computing its own time-evolution.” [Cha07]

The synthesis of Informational Structural Realism with the Computing Universe leads to Info-Computationalism [DC06] [DC10] [Mue10]. In short: information is the structure, the fabric of reality. The world exists independently from us (realist position of structural realism) in the form of proto-information. That proto-information becomes information for an agent in a process of interaction. Formalization of Info-Computational approach within Category Theory may be found in [Bur10].

Information and Computation in Biological and Intelligent Artificial Systems

Within the info-computational framework process of cognition is seen as a successive structuring (organizing) of data, where data are understood as simplest information units, signals acquired by an agent through the senses/sensors/instruments. Information as meaningful data, is turned into knowledge by a computational process going on in an agent. An agent is a physical system (living organism or an intelligent machine) or software possessing adaptive learning behaviors. Living agents have evolved from pre-biotic inorganic forms into increasingly complex systems able to self-organize, adapt to the environment and reproduce -all based on information processing.

Understanding information processing as computation [Bur05] implies that knowledge generation in biological agents involves natural computation, defined by [Mac04] as computation occurring in nature or inspired by that in nature. Part of natural computation characterized by the use of inexact solutions to computationally-hard tasks is called soft computing -many of bio-computing methods are of this soft character. It includes evolutionary computing, neurocomputing, fuzzy logic and probabilistic computing. [Zad98] emphasizes that an “essential aspect of soft computing is that its constituent methodologies are, for the most part, complementary and symbiotic rather than competitive and exclusive.” Natural computation is necessarily both symbolic and sub-symbolic information processing, which can be both discrete and continuous, as it corresponds to the dynamics of natural processes.

1. Informational Structures of Reality: Entities and Levels

In order to get a more specific understanding of info-computational thinking in multiscale information, granulation and aggregation, some definitions are in order.

Entity is used to mean anything that is considered to be discretely classifiable by some system. An entity can be composed of other entities (parts) -but it has its own identity. An entity could be a physical object or an abstract idea; within Informational Structural Realism both are different sorts of informational structures, which on the bottom level are data.

Granules are collections of information entities that usually originate at the numeric level (data-level) and are arranged together due to identity, similarity, functional or physical proximity, or alike.

Granular Computing is a new paradigm in which computing is understood as structured information processing in a hierarchy with multiple levels [Bar02]. Three different computational levels are distinguished which involve processing of information with increasing granularity: with the lowest level being numeric processing, the intermediate level processing larger information granules, and the highest level involving symbol-based processing. Within the info-computational framework we can continue to ever more extensive chunks of information that can be processed on increasingly higher levels: variable granulation (clustering/aggregation); system granulation (aggregation), concept granulation (component analysis) and so on. According to [Yao09], Granular Computing is a human-inspired approach, with important applications in the design and implementation of intelligent information systems. [Zad98] sees Granular Computing as the fundamental contribution to Fuzzy Logic which serves as a basis for the methodology of Computing with Words in which data sets have the form of propositions expressed in a natural language and words are interpreted as labels of granules. Zadeh goes as long as to suggesting Granular Logic as a hybrid of Fuzzy Logic with Granular Computing.

“In combination, the methodologies of soft computing, fuzzy logic, granular computing and computing with words are essential to the conception, design and utilization of information/intelligent systems because such systems reflect the pervasiveness of imprecision, uncertainty and partial truth in the world we live in.” [Zad98]

Information organization in levels: Conger [Con25] defined level as “a class of structures or processes which are distinguishable from others as being either higher or lower.” Levels are the consequence of two complementary processes of differentiation (separation, division, the result of analysis) and integration (the result of synthesis, which means they form as a result of grouping of some entities on a common “integrative levels”).

Levels reflect the inherent nature of reality (objective interpretation like levels of organization, levels of structure) but are always chosen among several possibilities, and in that way reflect our conceptualization of reality (subjective interpretation: levels of description, levels of representation, levels of analysis, levels of abstraction, [San04] and
Levels are used to organize informational granule for the reasons of simplicity and clarity. Level thinking is fundamental for representation, organization and synthesis of data, information, and knowledge. A structured organization of information seems to be one important way to outwit the limited capacity of human information processing, as shown by [Mil56].

Emergence is a phenomenon observed when entities on a higher level exhibit novel properties that are not observable on the lower levels. Levels correspond to aggregation of objects (informational granule) which are forming new wholes on a different scale of resolution. Life is an example of emergent phenomenon in objects (organisms) that have signaling and self-sustaining processes. Even though living organisms consist of inorganic parts, those parts show a qualitatively new behavior as a result of mutual interactions. Biology emerges from chemistry. Emmeche et al. [Stj97] identifies the following four primary levels of organization (integrative levels): physical, biological, psychological and sociological.

Emergence occurs in both space and time as “the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems.” [Go99].

Visual Cognition as Information Processing

Computational Visual Cognition is a study of human visual perceptual and cognitive abilities in a natural setting, including abilities at real world scene understanding, perception, recognition and memory alongside with the role of attention and learning. It uses tools and theories from image processing, cognitive science, computational vision, computer graphics and cognitive neuroscience. [Uli06] has developed computational models of vision that explain how humans recognize objects, perceive motion, use their visual world for task-relevant information, and create coherent representations of their environments. The main strategy here is learning about human visual cognition through construction of artifacts, using model-based reasoning.

Recent studies in biology, ethology and neuroscience have increased our knowledge of biological cognitive functions, and led to the understanding that the most important feature of cognition is its ability to deal efficiently with complexity, [GM95]. Insights into natural intelligence, together with the increase in power of computing machinery have brought us closer to the modeling of intelligent behavior. [DC08] argues that Info-Computationalism presents the most appropriate theoretical framework for understanding of the phenomena of cognition and intelligence in both biological and artificial systems. Cognition is seen as based on several levels of data processing [Min10] [Goe93] in a cognizing agent. Information processed from-sensory data processed into perception by an agent can be understood as an interface between the data (the world) and an agent’s perception of the world, [Ho09]. Patterns of information should thus be attributed both to the world and to the functions and structures of the brain. In an analogous way, knowledge can be understood as an interface between perception and cognition. Meaning and agency are the results of the information processing, and its refinement by relating to already existing (memorized) information. The meaning of an object in the external world is recognized through the process of perception of sensory signals, their processing through nervous system, comparison with memorized objects, and anticipation from memorized experiences.

Data, information, perceptual images and knowledge are organized in a multi-resolutional (multi-granular, multiscale) model of the brain and nervous system, [Min10]. Multi-resolutional representation has proven to be a good way of dealing with complexity in biological systems, and they are also being implemented in AI, [Goe93]. This is in accordance with Levels of Processing Models of memory [Cra72]. It should be noted that this one-directional sequential progression from shallow to deep processing, is only a first approximation, while in a more realistic model a combination of bottom-up processing (stimulus-driven) and top-down processing (concept driven) must be applied.

Computational Practices

“Artificial neural networks designed to implement even the most rudimentary forms of memory and knowledge extraction and adaptive behavior incorporate massively and symmetrically interconnected nodes; yet, in the cerebral cortex, the probability of a synaptic connection between any two arbitrarily chosen cells is on the order of 10^{-6}, i.e., so close to zero that a naive model might neglect this parameter altogether. The probability of a symmetric connection is even smaller (10^{-12}). How then, are thought and memory even possible? The solution appears to have been in the evolution of a modular, hierarchical cortical architecture, in which the modules are internally highly connected but only weakly interconnected with other modules. Appropriate inter-modular linkages are mediated indirectly via common linkages with higher level modules collectively known as association cortex.” (Emphasis added) [McN10].
Self-organization and self-description are fundamental intrinsic properties of natural intelligent systems. Learning is an essential part of those capabilities and it requires among others the development of a symbolic representations easy to maintain and use. In intelligent biological systems based upon a hierarchy of functional loops, each of these loops can be treated as a control system per se [Min10]. Generation of structures resulting from sensory processes and existing information in a nervous system are built in a multi-resolutional way, with many pattern recognition and control mechanisms hardwired.

Describing existing computing practices, [New82] introduces the concept of levels of analysis for a computer system, suggesting that there is a higher level than is usually addressed by computer scientists and calls that level the knowledge level, distinct from the symbolic level in which representation (logic) lies. Newell defines the knowledge level as embodying all the knowledge obtained or inferable from given information.

Similarly, in case of biological systems, Marr proposed that cognitive processes can be thought of as having three levels of description [Mar82] - an approach that can be generalized to any information processing system:

The top (computational) level describes the characteristics of the inputs and outputs of the computational problem solved by the process.

The middle (algorithmic) level provides the steps that transform the input into the output.

The bottom (implementational) level describes the physical machinery (hardware) that carries out the computation defined by algorithms.

Observe the difference between algorithmic level and computational level in Marr. In Church-Turing model of computation, those two coincide, and computation is identical with algorithm execution. The knowledge layer in Marr is a consequence of his thinking of computing as a real world process while Turing-Church computing is an idealization of computing as an abstract mathematical procedure.

In a similar vein, and focusing on Visual Cognition, [Dia07] finds that images contain only one sort of information - the perceptual (physical) information. Semantics comes from a human observer that perceives and interprets the image. Defining information based on Kolmogorov’s complexity and Chaitin’s algorithmic information, Diamant proposes a framework for visual information processing, which explicitly accounts for perceptual and cognitive image processing.

Understanding simulations and their underlying mechanisms is informative for several reasons, among others in order to understand our own mechanisms of grasping the reality, as human cognition in many ways resembles process of simulation and emulation. From the cognitive point of view, in order to conceptualize enormous flow of input data that bombards us through our sensory apparatus, we have to focus on limited amount of data organized in structures. Partly, as a result of evolution, this clusterization and aggregation of data happens in our nervous system automatically on the hardware level without our being aware of it (sub-symbolic information processing). From the results of neuroscience and cognitive science we are learning about the implementation of those computational mechanisms in living beings, which appear in a hierarchy of levels.

Simulation, Emulation, Aggregation

Over the last years, there have been several attempts to philosophically define computer simulation [Dur10]. Next we will briefly address two of such attempts that we consider the most prominent of the latest literature. We have picked them in the first place because they set the agenda for the contemporary philosophical discussion on computer simulations. The first one is due to Hartmann:

“Simulations are closely related to dynamic models. More concretely, a simulation results when the equations of the underlying dynamic model are solved. This model is designed to imitate the time-evolution of a real system. To put it another way, a simulation imitates one process by another process. In this definition, the term ‘process’ refers solely to some object or system whose state changes in time. If the simulation is run on a computer, it is called a computer simulation.” [Har96]

The mayor drawback in Hartmann’s definition is that it is too broad since it does not differentiate simulations from other similar activities such as imitation or emulation. As a matter of fact, he bases his definition of simulation on the concept of imitation, which he does not define. However simulation should be differentiated from imitation, since in the former case we expect to have knowledge of the target system whereas in the latter we do not. Something similar happens for the case of emulation.

The second definition of computer simulation is due to Humphreys. In this case, the author strengthens the definition of computer simulation by claiming that:

“A computer simulation is any computer-implemented method for exploring the properties of mathematical models where analytical methods are unavailable.” [Hum90]

Although Humphreys’ definition is more precise than Hartmann’s, it is still too general. It is not clear what he means by exploration of properties of mathematical models. Let us try to unfold this.

A major deficiency in Humphreys’ concept is that it restricts computer simulations to mathematical models, which is clearly not accurate enough for characterization. Computer simulations are based on algorithms, but an algorithm not necessarily has to be based on a mathematical model, for
it could be based on logical models as well as specifications of a system lacking of any mathematical or logical structure.

Several other definitions have been discussed in recent literature, but Hartmann and Humphreys have set the agenda for philosophical analysis of computer simulations.

The dilemma of computer simulations centers on the question of their value as tools for exploration of the real world. Many philosophers see the deficiency of simulation related to the materiality of experiments, which is not present in simulations. This claim is now known as the ‘materiality problem’ of computer simulations and has its standard conceptualization: “in genuine experiments, the same ‘material’ causes are at work in the experimental and target systems, while in simulations there is merely formal correspondence between the simulating and target systems […] inferences about target systems are more justified when experimental and target systems are made of the ‘same stuff’ than when they are made of different materials (as is the case in computer experiments)” [Par09].

Three solutions are in place: the lack of materiality of computer simulations make them defective as tools for production of relevant and reliable information (for example [Gua02]; [Mor05]; [Gie09]); the presence of materiality in experiments is ultimately, unimportant for purposes of investigations of real world phenomena (for instance [Mor09]; [Par09]; [Win09]), or some kind of middle point where the physicality of the computer, that is, the actual physical states in which the computer is switched when computing, serves as the basis for making the case of causality within the computer. There is a forth solution to this problem presented by Barberousse, Franceschelli and Imbert [Bar09] who claim that the representational and predictive power of computer simulations does not depend upon physical capacities, but upon a detailed analysis of their semantic levels.

In the following, definitions will be given in order to clarify ideas of simulation, emulation and aggregation.

*Emulation vs. Simulation:* “One system is said to emulate another when it performs in exactly the same way, though perhaps not at the same speed. A typical example would be emulation of one computer by (a program running on) another. You might use an emulation as a replacement for a system whereas you would use a simulation if you just wanted to analyze it and make predictions about it.” [The Free On-line Dictionary of Computing. © 1993-2000 1995-05-12]

*Aggregation* is the process of collecting together (aggregating) information from multiple sources. In order to increase semantic value of aggregation, a process of curation is often applied, which is done by humans who are making sense by synthesis and contextualization of aggregated information. The next step we may hope for is semantic web performing automatic intelligent aggregation and curation of information.

### Information Visualization

Information visualization is the interdisciplinary study of “the visual representation of large-scale collections of non-numerical information, such as files and lines of code in software systems, library and bibliographic databases, networks of relations on the internet, and so forth” [Fri08]. [Che10] gives the following more general definition: “Information visualization is concerned with the design, development, and application of computer generated interactive graphical representations of information.” In the more narrow sense, visualization deals with abstract, nonspatial or qualitative data which are transformed and represented so as to best impart meaning, using information design methods. Scientific Visualization on the other hand is more straightforward in a sense of interpretation as the data usually are numerical.

As [Che10] points out, the main goal of information visualization is for users to gain insights, so its role is to provide cognitive tools contributing to extended cognition.

The graphical representation of information is particularly suitable due to the massive hardware support for processing visual information which we have in the nervous system and especially in the brain. Once the simulation output has been constructed, the visual format is most easily deconstructed/decoded/translated by the brain into its internal representations because of the substantial hardware support for the visual information processing. In simulation models visualization is usually an important part, often accompanied by interactivity and animation (behavior of the system in time, and that is where it connects to visual cognition.

### Google Earth as an Example of Visualized Content Aggregation on Different Levels of Resolution/Organization of Content

Google Earth is a geographic information system presenting virtual globe which maps the Earth by the superimposition of satellite images, GIS 3D globe and aerial photography. It displays satellite images of varying resolution of the Earth’s surface, but most land is covered in at least 15 meters of resolution. One can overlay different layers of information like crime statistics, or information about culture, history and similar. Google Earth is using mashups with other web applications such as Wikipedia or Google News, which provides automatic aggregation of information using algorithms which carry out contextual analysis.

Google Earth is a good example of Granular Computing framework in which different levels are visually represented. On a low resolution satellite image one can observe large-scale phenomena such as weather patterns, and land relief of a globe, and in highest resolution image one views small-scale phenomena, such as the streets of a town and even particular houses.

The same principle of hierarchical information structur-
ing is used of all data representation in information processing systems, both artificial and biological ones. Biological agents, having necessarily limited resources, rely on layered data representation by grouping of chunks of information of different granularity. At different resolutions (granularities, levels of aggregation) different features and relationships emerge. Granular computing uses this property of cognition in designing more effective machine learning and reasoning systems that smoothly connect to human cognitive systems.

Conclusions

Being a young research field, (The first specialized international journal, Information Visualization, appeared in 2002). Information Visualization currently lacks adequate theoretical foundations. In search for theoretical foundations, theories and conceptual frameworks from other fields and disciplines are adopted, notably frameworks from human-computer interaction such as distributed cognition, [Sta08].

As defined by Shneiderman, Card and Mackinlay, “Information Visualization is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition”. Consequently, Information Visualization designs can be seen as tools - cognitive extensions based on active vision which uses graphic designs as cognitive amplifiers.

We present an info-computational theoretical approach unifying the principles governing content aggregation, information visualization and emergent properties in computer simulations and cognition. It puts into a common context information granulation/aggregation, granular computing paradigm, level method of information organization, emergent properties and info-computational character of cognition as an information processing mechanism that in many respects resembles simulation and emulation processes.

References


[De06] Dodig-Crnkovic G.: Investigations into information semantics and ethics of computing, 2006. 2
[Hum90] Humphreys P.: Computer simulations. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association (1990), 497–506. 4
Marr D.: Vision: a computational investigation into the human representation and processing of visual information. W. H. Freeman, 1982. 4

McNaughton B. L.: Cortical hierarchies, sleep, and the extraction of knowledge from memory. Artificial Intelligence 174, 2 (2010), 205–214. 3

Miller G. A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63 (1956), 81–97. 3


Morrison M.: Models, measurement and computer simulation: the changing face of experimentation. Philosophical Studies 143 (2009), 33–47. 5


Needham J.: Herbert Spencer Lecture. Time, the refreshing river. 1943. 3


Parker W.: Does matter really matter? computer simulations, experiments, and materiality. Synthese 169, 3 (2009), 483–496. 5


Vedral V.: Decoding Reality: The Universe as Quantum Information. Oxford University Press, 2010. 1


