

Deciphering Aging

Abstract

Denigma is a Digital Decipher Machine that can encode and decode data in the web. It is specifically devoted to decipher the aging process by utilizing and systematizing the rapidly increasing amount of available biological data and any other information with relevance to aging research. Denigma structures the information in the form of facts and enables modular access to diverse data and logical reasoning on the constructed knowledge base. This enables decoding algorithms to perform global inference and planning on the web-scale and therefore enables the identification of effective interventions by considering all the previous knowledge and making highly confident and testable predictions.

Introduction

Aging is an enormous problem that needs a massive concerted effort to be addressed. Such a project will not be possible without establishing an extensive information technological platform that will be used to unify data, provide collaborative tools for researchers and longevity activists and automate the discovery of novel therapeutics to be tested.

That is why the International Longevity Alliance (ILA) initiated the construction of a digital decipher machine, Denigma [<http://denigma.de>], to decode the aging process, make aging research more collaborative and efficient, and facilitate the search for powerful means to ameliorate the degenerative aging process.

Denigma functions as ILA's main information technological platform both for communication and scientific purposes. In the long range, the potential life-extending interventions that may be indicated by data-mining using the Denigma, will need to be further tested in real life in specially established or expanded existing testing center(s) [Figure 1].

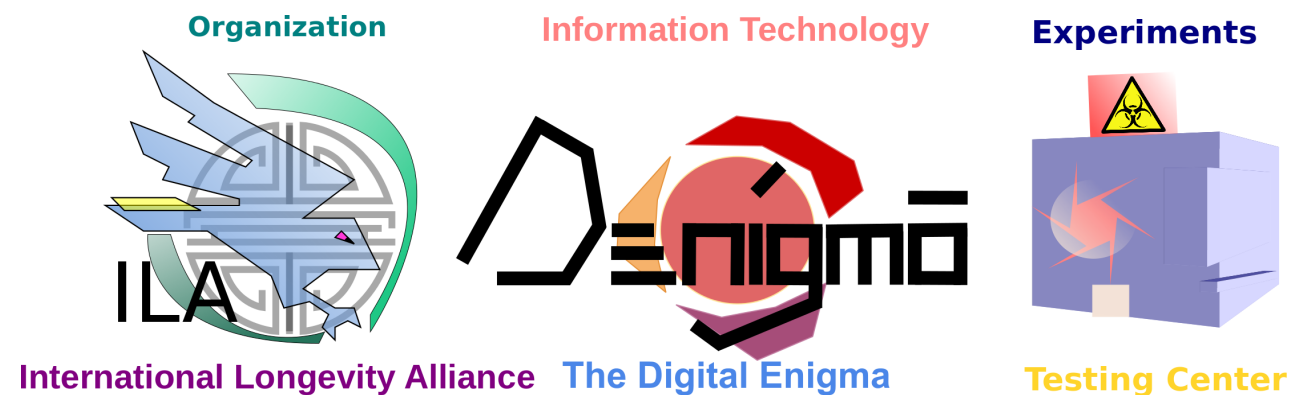


Figure 1: The Vision.

The construction of a digital decipher machine will allow the creation of a web intelligence capable of reverse-engineering the aging process. Specifically Denigma uses advanced information technologies such as semantic web and machine learning bundled with information theory and combined with crowdsourcing for longevity research. Its development is lead by a group of enthusiasts consisting of aging research scientists and software development engineers, as well as social networkers, who all together form the Global Computing Initiative (GCI).

Data Unification for Deciphering Aging

Currently biomedical knowledge is scattered in heterogeneous formats. It is present in the form of unstructured academic articles, semi-structured datasets, structured databases and non-accessible in the minds of experts. This renders such heterogeneous knowledge inaccessible for global computing [Attwood et al., 2009].

In order to provide a holistic view for researchers and run various machine learning algorithms to infer new knowledge, the data should be represented in a well-structured format. This format should be at the same time both machine and human readable. For this we need a knowledge graph and the semantic web provides the most suitable technology, functioning as a widely accepted standard and initially developed to create a new, knowledge-based generation of the web.

A key technology to enable such a multi-layered data integration is ontology engineering. An ontology is an explicit formal specification of a shared conceptualization [Gruber T, 1995]. It is the basis for the so-called Semantic Web Stack, a set of technologies that enables researchers to describe heterogeneous data and run reasoning computer algorithms on the data.

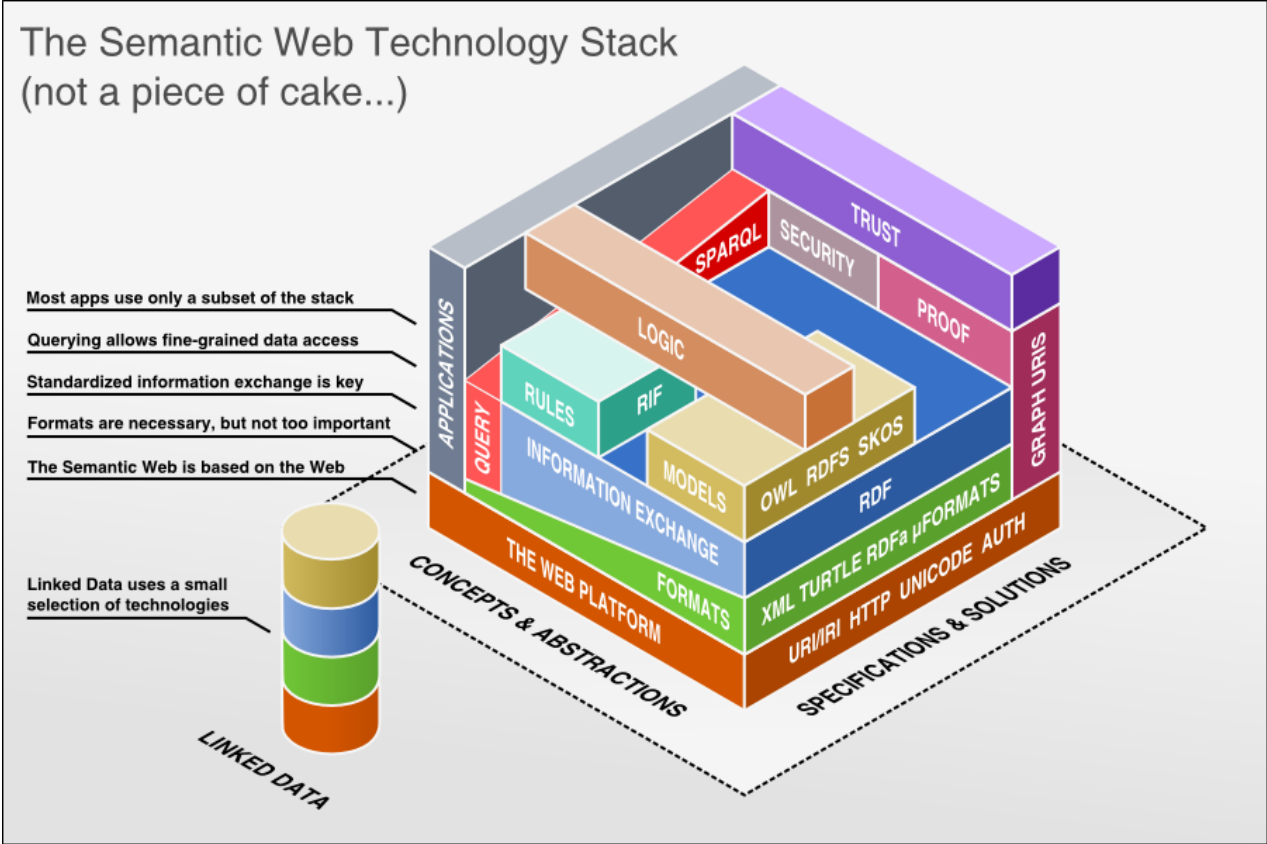


Figure 2: Semantic Web Stack.

Open Distributed Science for Aging Research

In order to create a semantic graph, the knowledge from research papers in the form of free text needs to be converted into a structured form, while individual articles, datasets and databases need to be mapped for modular access and converted into the same structure. Ontologies specific for aging research are necessary and they are being developed on Denigma. A collective effort is needed to establish the initial structure of the data. Data annotation and ontology creation processes can be automated, but only partially and therefore require human reasoning.

Linked data allows to combine machine learning with the crowdsourcing approach where many tasks that need "human computations" are effectively outsourced to the crowd. Our crowdsourcing infrastructure is made to attract longevity activists to help scientists, complement and improve machine learning where humans do what machines cannot do well.

Denigma itself is the result of a crowdsourcing effort of researchers of aging and longevity as well as scientists from other fields.

We work in strong collaboration with leading researchers on aging such as João Pedro de Magalhães from the Institute of Integrative Biology (Liverpool, UK), the pioneer of the genomics of aging and of the application of systems biology for aging research; and Aubrey de Grey, head of the SENS (Strategies for Engineered Negligible Senescence) Research Foundation (US/UK), and the pioneer of tackling aging as an engineering problem and of the use of regenerative medicine to reverse aging, among many other scientists.

We have been outsourcing tasks to and are mentoring students of the SENS Research Education.

Further we develop novel ways of making the aging process visually more attractive, so as to engage citizen scientists and utilize the power of the crowd to complement automated algorithms to knowledge base creation, as well as educate the public about the problem of aging.

There are numerous other applications developed on Denigma that are valuable for aging research, such as a semantic management system that harbours information about experts researching on aging, transhumanists and life extension activists, as well as functionalities to enable real-time collaborations. Specifically we develop collaborative features to ease cooperation for researchers and activists.

Constructing a Knowledge Base on Aging

In order to decipher aging, we need to have a comprehensive and consistent knowledge base.

Aging research is chaotic and not well developed, compared to other disciplines of biology and other branches of natural sciences. Therefore, even the simplest data unification together with a good user interface and basic machine learning can have a large impact on the field.

So far Denigma already has gathered a tremendous amount of facts about aging research. Thus, it is easily possible to see what has already been achieved in anti-aging experiments and which factors influence aging.

Specifically Denigma creates a novel form of the semantic network based on Data Units that is compatible with the existing technologies, but more flexible [Figure 3].

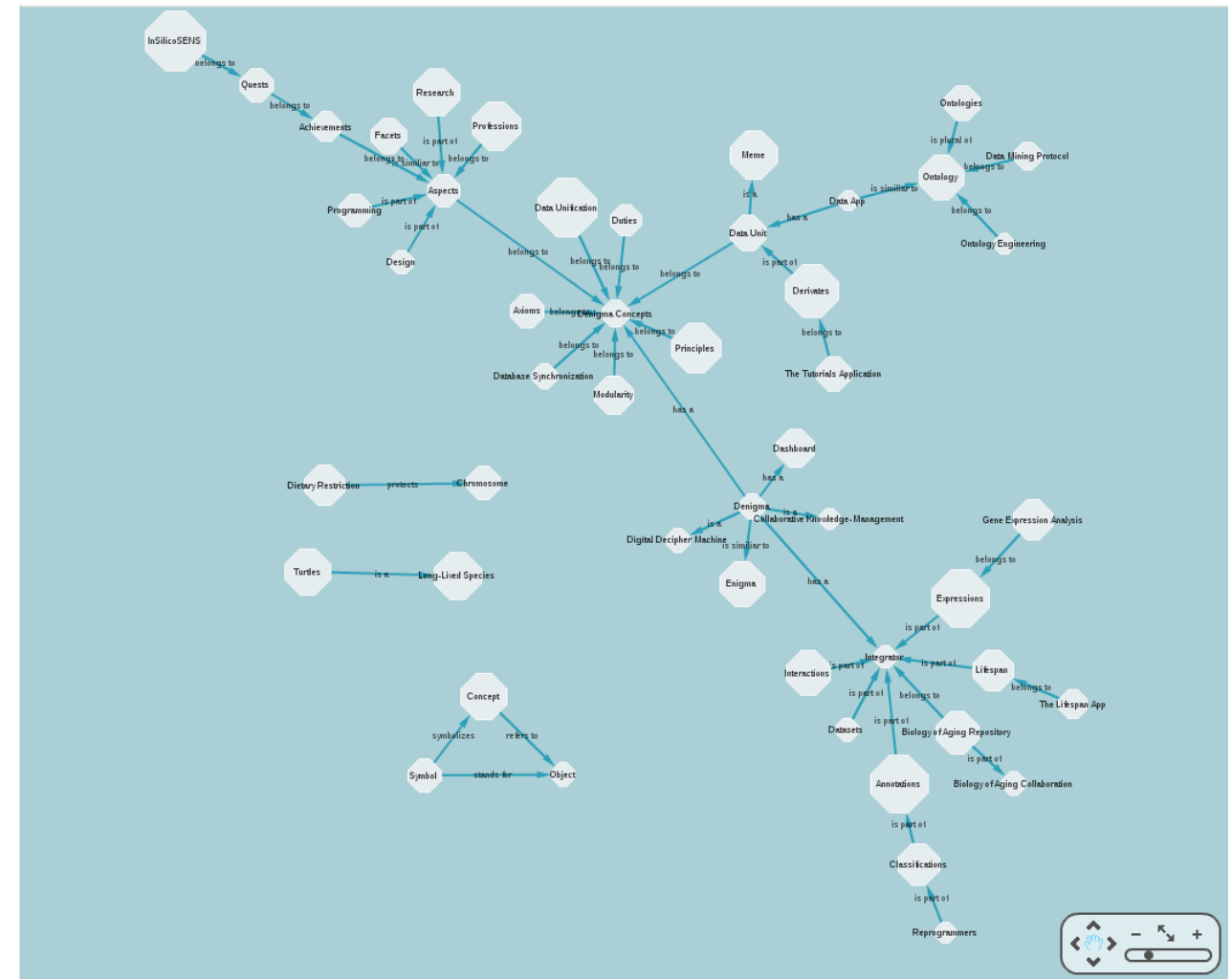


Figure 3: Data Units.

The Lifespan Application

The heart of Denigma is the Lifespan application which systematizes data from lifespan studies, ranging from lifespan experiments, through life extending interventions to individual factors that limit the lifespan.

As the main goal of aging research is finding ways of ameliorating aging for humans, there is a need to conduct "lifespan intervention" experiments on animals to extend their lifespan, and then infer from that data for humans [Figure 4].

The Lifespan application systematically maps the determinants of the lifespan, collected from many resources, ranging from studies that report lifespan experiments, through personal measurements to individual factors that influence the aging process.

The resources include:

- Studies
- Experiments
- Measurements
- Comparisons
- Interventions
- Factors

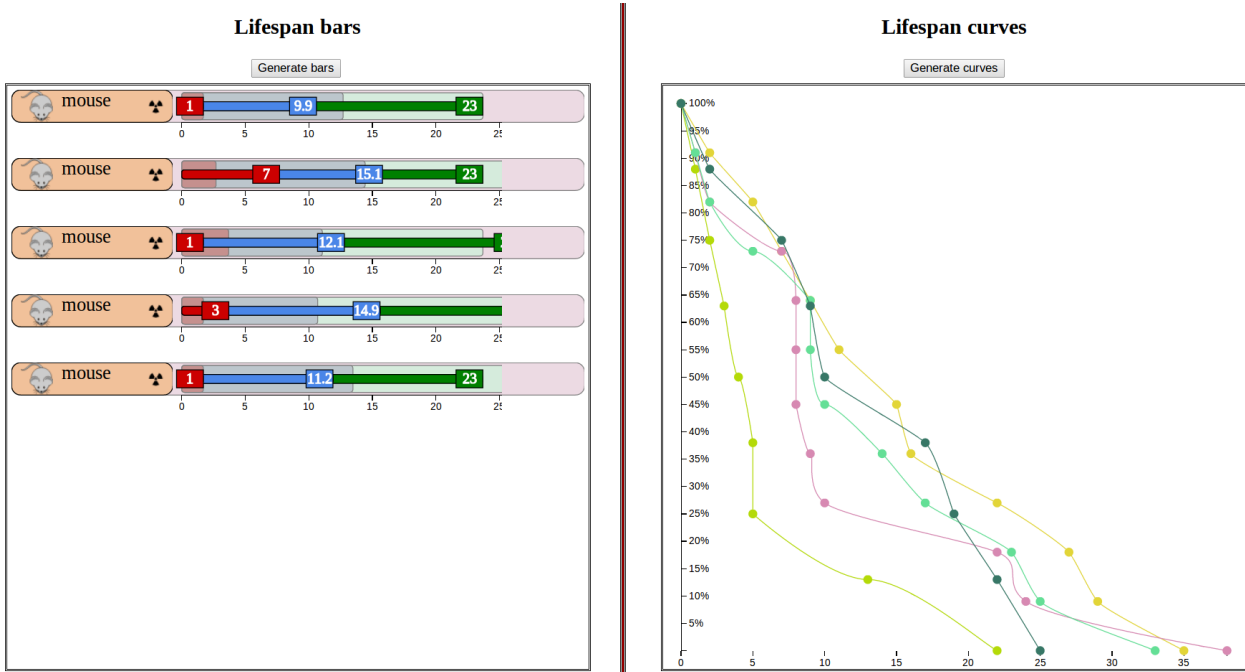


Figure 4: Lifespan Experiments Charts.

We recently started the construction of the world's most comprehensive database on the genetic variants associated with longevity in humans, which will help computer-aided targeted drug discovery to be used to extend the human healthspan.

The Gene Application

The aging process is very plastic. It can be modulated by genetic as well as environmental factors. Single gene mutations identified in various model organisms can greatly extend the lifespan, by up to 10 fold [Shmookler Reis et al. 2009]. Importantly, it appears that most of these genes are highly conserved between species [Wuttke et al. 2012].

We utilize functional genomics information, such as gene expression activity changes and interactions between biological entities to infer new possible factors capable of extending the lifespan. Network analyses of genes interactions have been conducted by us [Figure 5].

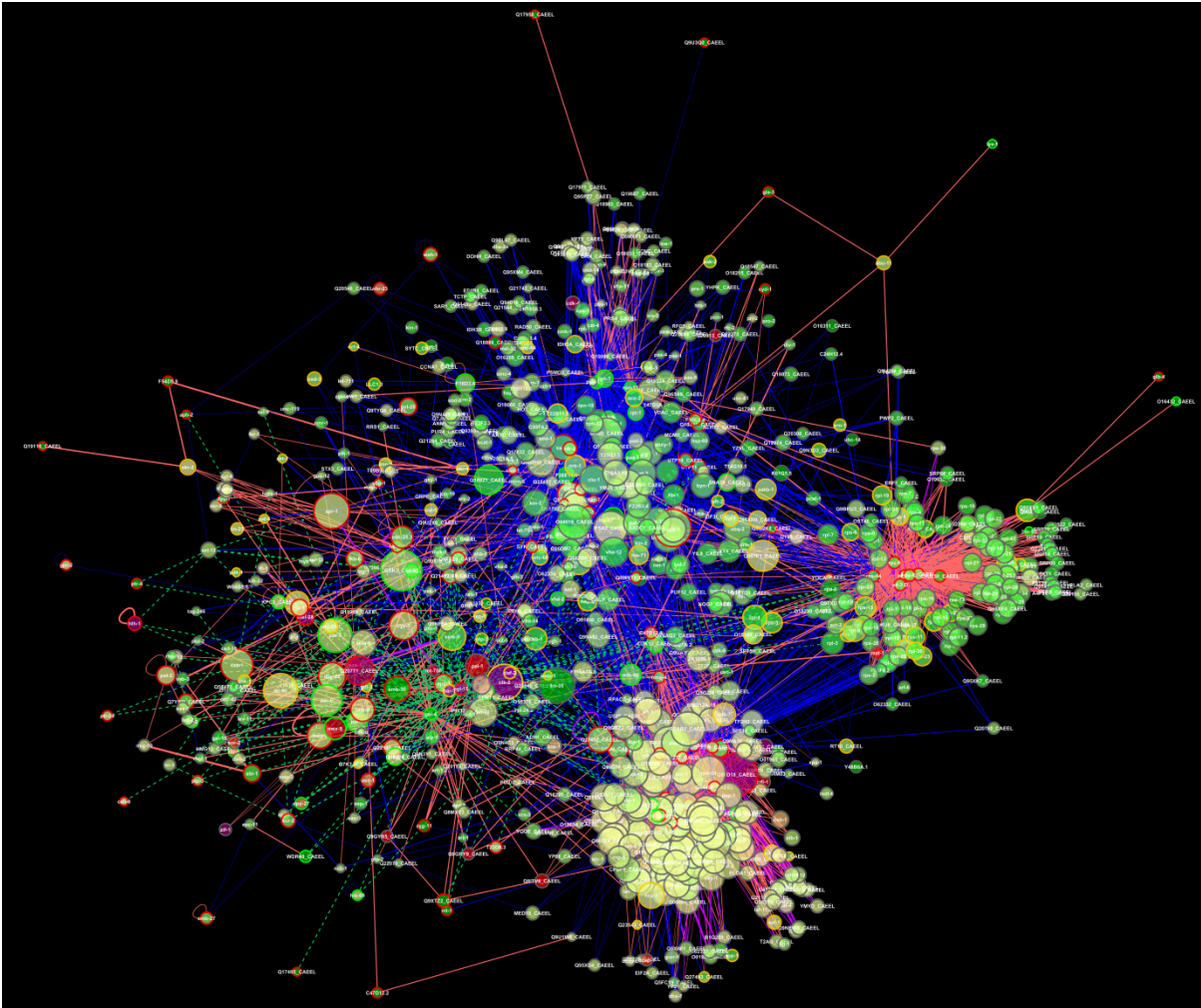


Figure 5: Interaction Network of Aging Genes.

Finding Effective Anti-Aging Interventions

Computational methods developed on Denigma will enable researchers to identify effective therapeutics. Novel interventions can be found by utilizing top-down approaches via the use of high-throughput *omics* data or bottom-up targeted approaches, including:

- Gene expression pattern matching
- Network-based dissection
- Computer-aided drug discovery
- Mutual information to predict efficiency
- Fact-based reasoning with ontologies

Gene Expression Pattern Matching

Gene expression activity patterns from transcriptomics data of already known lifespan extending interventions are used to derive common molecular signatures associated with longevity. Those signatures are then matched with gene expression profiles of drugs that trigger similar gene expression patterns. Drugs that induce similar gene expression changes are with high accuracy lifespan-extending drugs. Similar meta-analysis of aging-related gene expression changes is used to derive a common molecular signature associated with aging. The comparison of this signature with the gene expression profiles of drugs allows the researchers to identify those components that reverse the age-related gene expression changes. Those drugs that reverse aging gene expression changes can be powerful anti-aging drugs capable of reversing aging.

Network-Base Dissection

Having a comprehensive inventory of all the factors that are associated with lifespan determination, enables to apply network-approaches by utilizing interactomics data. For instance, given a list of the genes associated with the process of aging, the researcher can apply the "guilt-by-association" concept which basically states that genes with more interactions than expected by chance with genes associated with the given process are likely to also play a role in that process, i.e. are potential drug targets.

Computer-Aided Drug Discovery

By leveraging distributed computing algorithms, we can identify drugs that fit exactly into the three-dimensional structure of specific gene products, for example, products of aging genes. This approach allows to develop drugs with specific effects.

Mutual Information to Predict Efficiency

As mutual information provides the exact estimate of similarity between various model systems, it will be possible to predict the efficacy of a yet untested drug or treatment using the estimates of its similarity (mutual information) with other tested drugs and treatments along with the similarity of model systems to which they are applied.

Fact-Based Reasoning with Ontologies

Logical reasoning of a consistence knowledge base with appropriately designed ontologies allows to infer the causality of the chain of age-related changes and can suggest, via revealing implicit hidden knowledge, which therapies are most effective to reverse the drivers of age-related changes.

Therefore, by utilizing such computational approaches, the Denigma will be able to suggest experiments to provide insights into the mechanism of aging and its potential reversal, taking into account the current existing knowledge and data, and to provide methods to estimate the efficiency of potential life-extending interventions.

An Ecosystem of Reusable Applications

In fact information technology is a field where a group of several developers working full-time can change a lot, as features that we develop will increase the productivity of many researchers involved in the field. And as for any IT platform, all features of Denigma can also be used in many other fields (both research-related and unrelated) that will provide additional opportunities.

We are developing tools and applications which are primarily used for research on longevity. We are also training the crowd in programming to facilitate research and development in the field.

Some of the software products that we produce are of general usability, while other kinds are domain specific. The software that we develop has the potential to become commercialized.

Applicability to other Domains:

- "Denigma for other fields". Many applications that we develop for aging can be reused for other fields.
- Knowledge management platforms. The current knowledge management platforms (like semantic wiki) that are used by corporations and academia were developed long time ago and do not meet the challenges of today. So the features that we develop for aging researchers and activists will be also applicable in those areas.
- Community building possibilities. Lifespan-extensionist communities can be strengthened and recruited using the Denigma communication services.
- Educational opportunities. We develop graphs and other methods of visualization as well as quests that, in combination with other features including gamification, can be easily used for education purposes anywhere.
- Crowdsourcing for research. We are working on features that let researchers outsource parts of their work to volunteers. Projects like "Aging Genes" progress well with this approach.
- Further benefits. When creating an ecosystem, it is often not known beforehand what opportunities can emerge. Yet, what is important in the Denigma platform is that here technologies/features exist together with the user community that utilizes them. Such a combination always provides new opportunities that were not known beforehand. Denigma is open-source which makes it easier to expand and attract new people. In the same time, it is possible to provide some paid services and consultancy on top of the open-source Denigma platform.

Conclusion

Denigma provides uniquely robust and flexible features for lifespan extension research and development. It is at the intersection of science, collaboration and machine learning [Figure 6]. It enhances scientific discoveries by combining human and machine intelligence, via supporting collaborations and utilizing crowdsourcing with artificial intelligence and machine learning.

Nothing is perfect from the beginning, besides the concept of Denigma. Denigma is as strong as diamond and as flexible as carbon-fiber, but in it's very heart it's all silicon.

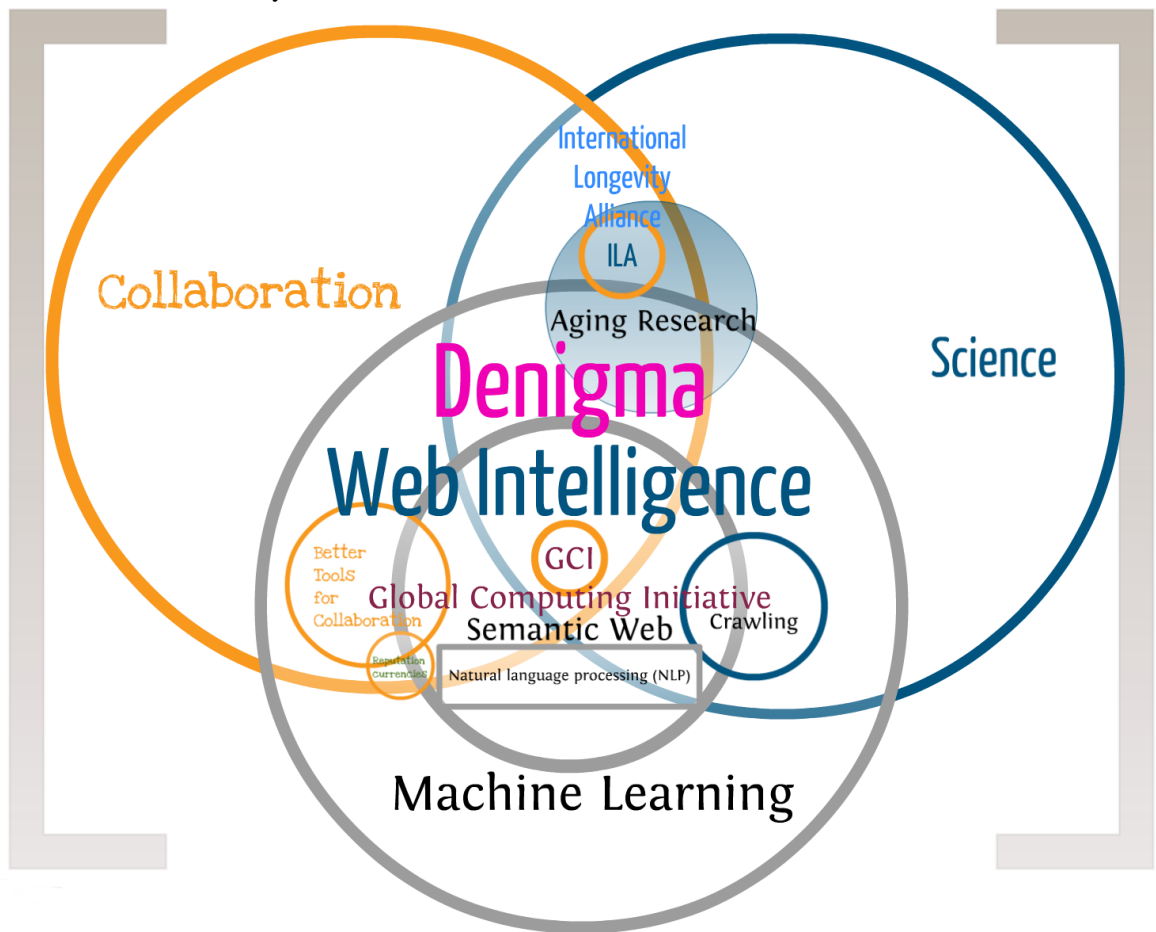


Figure 6: Denigma Concept.

References

Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, Thorne D (2009) Calling International Rescue: knowledge lost in literature and data landslide! The Biochemical journal 424: 317-33.

Gruber T (1995) Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5/6):907–928.

Shmookler Reis RJ, Bharill P, Tazearslan C, Ayyadevara S (2009) Extreme-longevity mutations orchestrate silencing of multiple signaling pathways. Biochimica et biophysica acta 1790: 1075-83.

Wuttke D, Connor R, Vora C, Craig T, Li Y, Wood S, Vasieva O, Shmookler Reis R, Tang F, de Magalhães JP (2012) Dissecting the gene network of dietary restriction to identify evolutionarily conserved pathways and new functional genes. PLoS genetics 8: e1002834.