HENUFOOD: Development of New Methodologies and Emergent Technologies for Showing Food with Health Claims on Chronic Diseases Risk Reduction in the Middle Age of Life

G. Anzaldi, X. Domingo, A. Moreno and P. De La Peña

Abstract HENUFOOD looks forward reduce chronic disease pathologies risk factor and, in this way, improve adult population health, between the range of 45 and 65 years. However, the benefits of this project, based on healthy ingredients and foods development, try to reach the rest of the population from the beginning up to the seniors. The main objective of HENUFOOD is discovering the healthy benefits from aliments using innovative methodologies, and scientifically demonstrate it. That will permit develop value products at nutritional level and demonstrate their health effects. These foods must keep on being foods, and must demonstrate their effects in quantities which are usually consumed in a diet. The project is looking forward determining in a clear way which foods or ingredients are absorbed by the organism and produce the beneficial effect that they are supposed to. This paper will focus on describing the ICT platform developed to support the scientists reach that purpose.

G. Anzaldi (✉) · X. Domingo
Barcelona Digital Technology Centre, Lleida, Catalonia, Spain
e-mail: ganzaldi@bdigital.org

X. Domingo
e-mail: xdomingo@bdigital.org

A. Moreno · P. De La Peña
Ibermática, Donostia, Basque country, Spain
e-mail: ai.moreno@ibermatica.com

P. De La Peña
e-mail: pm.delapena@ibermatica.com

I. González Alonso (ed.), *International Technology Robotics Applications*,
Intelligent Systems, Control and Automation: Science and Engineering 70,
DOI: 10.1007/978-3-319-02332-8_2, © Springer International Publishing Switzerland 2014

1 Introduction

Health is one of the main concerns of developed societies. However, the pace of modern life has come to people leaving the healthy eating habits which are a key factor to prevent the incidence of chronic diseases. As said by the WHO [1], chronic diseases are increasing and in 2020 will be responsible of 75 % of total world deaths. However, fighting against the risk of having non transmittable chronic diseases is on our hand. The experts agree that when people follow a balanced diet in a continued period, the organism acquires a greater response capacity in front of possible health incidences. With this aim, food industries have started to offer products with nutrition or health claims which, in addition, will help them to improve their competitiveness in the market.

Moreover in accordance with Regulation (EC) 1924/2006 of the European Council and Parliament [2] of 20 December 2006 of nutrition and health properties of food, economic operators can only use nutrition or health claims on their products if and only if, such declaration is based on scientific evidence and, in addition, it is understandable to the average consumer.

The 15th article of Regulation (EC) 1924/2006 of the European Council and Parliament, states which should be the studies to be submitted to applications: the studies and other material mentioned in the 15th article, 3rd paragraph, 'c' and 'e' points. The studies will mainly consist of human studies and, in the case of statements regarding the development and health of children, children studies. These studies will be presented based on a hierarchy depending on the study design. This hierarchy will reflect the relative weights of the evidence that may be obtained. The outcomes of that studies should be clearly explained to target markets, and for this reason, some of the information to be clearly explained includes: (a) the target population for the intended health claim, (b) the amount of the nutrient or other substance, or food or food category, and pattern of consumption required to obtain the claimed beneficial effect, (c) where appropriate, a statement addressed to persons who should avoid using the nutrient or other substance, or the food or food category, which is the subject of the health claim, (d) a warning for the nutrient or other substance, or the food or food category, which may pose a health risk if consumed to excess, and (e) any other restrictions of use and preparation or use instructions.

Thus, with the current European legislation on the hand, all the foods with healthy or functional properties must scientifically demonstrate that they result in a beneficial effect in one or more organism functions, apart from the intrinsic nutritive effects, in a way that they result to improve the health and wellness, decrease disease risk, or the two things together. These foods must keep on being foods, and must demonstrate their effects in quantities which are usually consumed in a diet. So, they cannot be capsules but foods that are part of our daily diet. The project is looking forward determining in a clear way which foods or ingredients are absorbed by the organism and produce the beneficial effect that they are supposed to. Concretely, HENUFOOD will research on fiber, fatty acids and

phytosterols, probiotics, active peptides, photochemical and calcium, in order to detect the beneficial effects that they have in some health areas: cardiovascular protection, digestive and cholesterol metabolism, antioxidant action, immune homeostasis, comfort and intestinal health, glucose homeostasis and weight control, osseous and articular system.

To evidence this health benefits and scientifically demonstrate them, controlled clinical trials will be used. Controlled clinical trials are studies done by scientists on a concrete domain with the objective of execute a prospective analysis that attempts to compare the effect and value of one or more interventions in humans under specific conditions. To support these activities, a specific platform will be built. This platform will support scientists on their everyday tasks including researching the state of the art of their concrete study, advising and supervising the correct clinical trial execution, helping in the interpretation of the results, running analytical processes, semantically understanding the necessities by a user friendly interface based on query answering techniques, and acting as a decision support system. This paper will focus on that ICT platform intended for support the general objectives of HENUFOOD project.

2 Objectives

Taking in consideration the previously mentioned regulation context and industries necessities, HENUFOOD aims to discover and scientifically support the healthy benefits from aliments using innovative methodologies. These innovative methodologies include "omics" (transcriptomics, proteomics, metabolomics…) and the image diagnose (Functional Nuclear Magnetic Resonance-FNMR), which will let studying new biomarkers and analyze the changes that the organism suffers after the ingestion of some foods or ingredients. In this way, it is expected to guide the food development and production to an optimal nutrition, understood as the consumption of foods that provide clear and proven benefits for the risk of chronic diseases prevention.

From our concrete point of interest, our technical objective is to develop a general computational platform which will support evidence, outcomes validation, and decision support applied to procedures and scientific trials which are related to the research of beneficial food properties and their impact on health. The required functionalities to be covered by HENUFOOD technological objectives are:

- Integrate data from different biological models and generate conclusions.
- Optimize the design of nutritional intervention studies in humans.
- Procedures and automatisms helping in understanding the action mechanism. based on the various results in the test samples.
- Build a health risk prevention system.

- Application of computational methods to predict, using computer simulation, healthy effects of nutrients, as well as obtaining statistical justifications for such actions.

The activities need to build a unified platform which meets these functionalities and, in addition, offers management support and scientific trials conclusion guidance, validation and outcome management. Three separated but intrinsically related stages are considered:

1. Obtaining, pre-processing and treatment of data.
2. Automatic models validation, intelligent decision support and scientific support of the validated models.
3. Intelligent and intuitive data interpretation.

3 Impact and Benefit

The biomarkers that will be studied are focused on people health problems that are in the middle age of life: digestive and sensory changes associated with age, osteoarthritis, neurobiological changes associated with aging, circulatory disorders peripheral arterial hypertension, metabolic syndrome…

The selection of this group of people is because, at present, various WHO [1] studies have confirmed that in the coming decades, both health related no transmittable food chronic diseases (obesity, diabetes and cardiovascular) derived from the current way of life, and the increase of health costs to be supported by the developed and developing countries to address these diseases, are expected to increase worldwide. In this context, it should be noted that food and eating habits are the main non-genetic aspects related to health, which have an important role in the prevention of these facts and resources saving.

HENUFOOD looks forward reduce chronic disease pathologies risk factor between the range of 45 and 65 years. However, the benefits of this project, based on healthy ingredients and foods development, try to reach the rest of the population from the beginning up to the seniors.

4 HENUFOOD Platform

HENUFOOD solution is based on a central knowledge base which contains all the information needed for all the implied processes in the platform. This information includes domain ontology with medical and nutritional related terms, state of the art of the corresponding research groups and their publications, new knowledge acquired during the usage of the tools, and the relevant results of running analytical and data mining processes over the data. This ontology is the base of four differentiated modules:

- The bibliographic and state of the art research module: This module is responsible of collect information from different data sources. Natural language processing techniques [3, 4] are used together with content indexing. Then Knowledge Discovery in Databases [5–7] processes are applied and relevant data is added to the ontology. More information about this can be found at [8].
- The clinical trial management tool: This module is responsible of allowing scientists introduce different data from different controlled clinical trials and their later information treatment. It is also able to show alerts when necessary, in example, when CONSORT [9] recommendations are not accomplished, or when the concrete sample which is being introduced shows notorious differences in its characteristics than the other. The latter fact that could suggest an error in the sample collection or manipulation.
- The analytical wizard: This module is responsible of analyzing the information and running the corresponding statistical or data mining processes to the data. The interactions are done using a wizard based application which guides the user throughout the process. When new relevant knowledge is discovered, it is added to the central ontology. Examples of used algorithms are FP-Growth [10], A-priori [11], and other association, classification, correlation, prediction or segmentation algorithms.
- The avatar graphical user interface: This module is responsible of interact with scientists as if it was a personal assistant. The basic idea under this development is to allow scientists to ask for concrete analysis or questions over the state of the art in natural language. The system is able to answer to these questions in natural language too, or even run some processes or take some actions, depending on the context.

5 Work Done

This paper will explain the advances achieved related to the objectives of the platform. This section will be divided in different parts according to the stages described on the previous section.

5.1 Obtaining, Pre-processing and Treatment of Data

The comprehensive platform, which aims to solve the problems associated with intelligent information management, is focusing in two aspects regarding the research on methods of obtaining information: the indicator results obtained by biomarkers on the one hand, and on the other hand, the integration of unstructured data from external data sources (intranet, extranet, specialized databases and images).

In terms of the compilation of active substances, natural ingredients and its relationship with health, the medicine based on evidence paradigm has been taken as the basis. With that premises, relevant and innovative knowledge has been acquired by accessing semantically to the Linked Data cloud. In addition, multiple connectors to data sources related to the bio-nutritional state of the art have been developed. A bibliographic aggregation platform over the annotations base in natural language and the information extraction based on the platform ontology has been built.

Together with the previous functionality, an automatic extraction system has been developed to obtain relevant knowledge according to the data provided by the trial samples set, so that, depending on targets, decide which models to be applied (statistical models and data mining), and which are the main and multivariable attributes to feed that models. This automatic extraction system is based on a wizard guidance, which automates the more complex processes within CRISP-DM [12] methodology for managing the data mining information stream. These previously functionalities can be checked in a summarized schema in Fig. 1.

On the other side, focusing on the automatic evaluation of medical images, taking as a reference the automatic extraction of interest zones, graphical targets have been defined over the generality of the image. To accomplish that, data mining algorithms have been used, especially SVM and neuronal networks. However, traditional artificial vision matching techniques have been also applied. To improve the user interaction with these applications, a unified visual analysis solution with both technologies embedded has been designed.

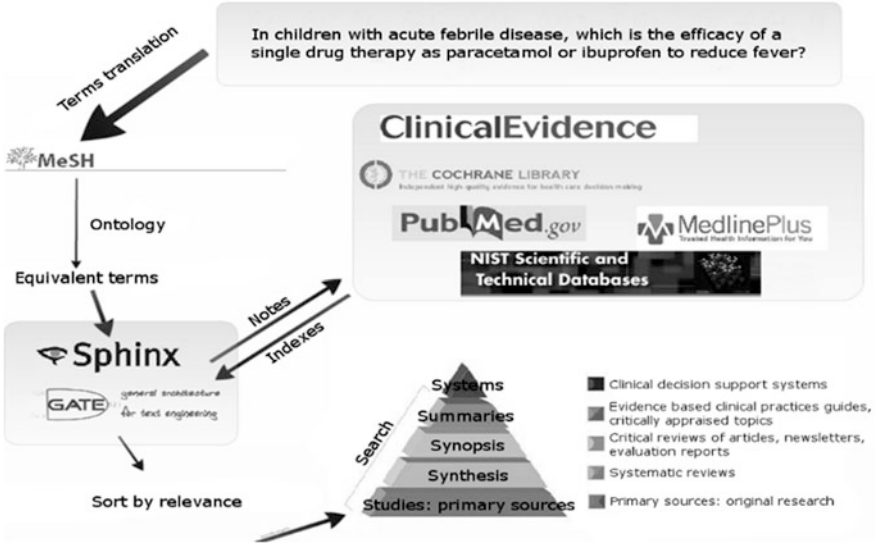

Fig. 1 Obtaining, pre-processing and treatment of data

5.2 Automatic Models Validation, Intelligent Decision Support and Scientific Support of the Validated Models

In parallel to the functionalities previously described, a modelling system has been developed. This system is using different state of the art techniques, such as profile and modelling comparison (multivariate analysis): ANOVA, PCA, PLSDA, and others, together with an association rules extraction system and a semantic inference engine (SWRL) over that information. This system, taking in consideration the study results, is able to determine the model convergence, the confidence level based on the original premises (analysis of the null hypothesis) and also to explain in a contextual way the outcomes. The system is also able to predict and respond how the system would respond towards concrete changes in input indicators.

This modelling system offers an intelligent support for clinical trials modelling and their simulation. It is empowered with a rule based expert system which will be responsible of checking in real time the rules and knowledge obtained in the previous subsection, together with the manual introduced ones. Some of the introduced rules are the ones coming from the CONSORT statements or the good clinical practices. Simplified architecture of these systems can be seen in Fig. 2.

To semantically support the entire platform, a central ontology is used. This ontology is the main kernel of the decision and control system and, in addition, it allows using a common vocabulary, integrate internal information, disseminate the trials outcomes, control the trial lifecycle, and validate, taking in consideration predefined quality rules, the outcomes, the samples and their priori defined objectives.

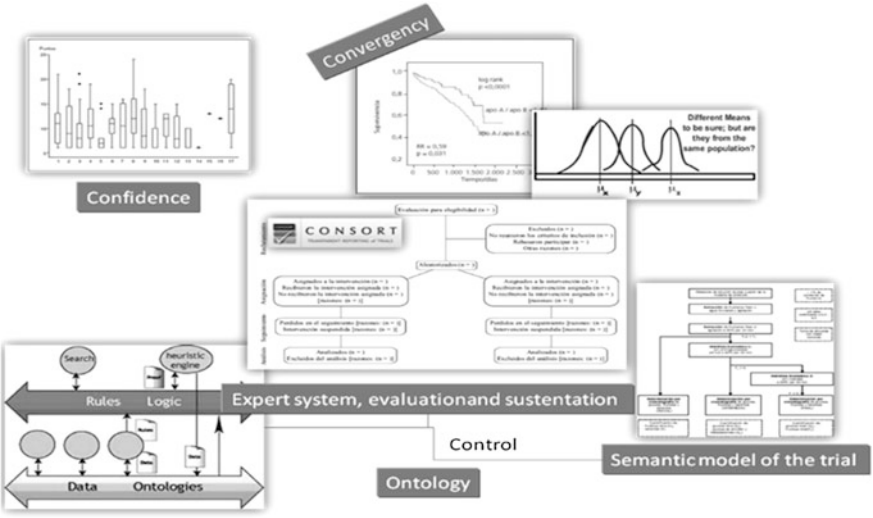

Fig. 2 Automatic models validation, intelligent decision support and scientific support of the validated models

This ontology includes and is based on other taxonomies as SNOMED [13], and pretends to match all the natural language described information with UMLS codification to later populate it. It also includes all the necessary terminology to successfully represent the concepts involved in the support and validation control of the target models. Additionally, the outcomes obtained using data mining processes are also included when relevant.

5.3 Intelligent and Intuitive Data Interpretation

The data accessing and interpretation, has to be done in an intuitive way. In this way, the interaction with the decision support system is done using questions in natural language. The first steps of this functionality include an interaction with the data mining and expert systems using direct questions over the object of study. Examples are "which are the independent variables?", "which are the relevant attributes?", "which are the recommended methods to apply?" or "which are the restrictive conditions in the models?". In fact, it is a controlled language which is abstracted from data base and the data mining terminology to allow high level interaction.

To make this interaction more accessible, an avatar, with which the interaction is actually done, has been designed. Thanks to that, scientist can ask questions or ask for some processes to be run and the avatar responds according to the context using natural language, running some statistical or data mining processes or similar as required.

Thus, the complete process of information compilation and manipulation, the knowledge extraction, the trial execution support and control, the analytical processes parameterization and execution, and finally, the results interpretation are executed in a transparent way to the scientists, as seen in Fig. 3. A general example of use of the platform will comprise the following steps:

- Scientist has an initial hypothesis and research the state of the art for similar studies and outcomes. The system answers with related literature and previous outcomes obtained. In case outcomes are significantly negatives, the system suggests trying another method or study. The avatar is the primary interface used with that aim.
- During trial execution, the system checks whether the partial results are aligned with the initial hypothesis, showing alarms when necessary. It also checks the accomplishment of regulations and good practices.
- After trial execution, the system runs analytical processes to check whether the hypothesis has been confirmed on the one hand, and on the other hand, discover new relevant non-obvious knowledge. The results are also introduced on the knowledge base.
- During all process, the system checks automatically the data sources for new relevant information to annotate it in the knowledge base.

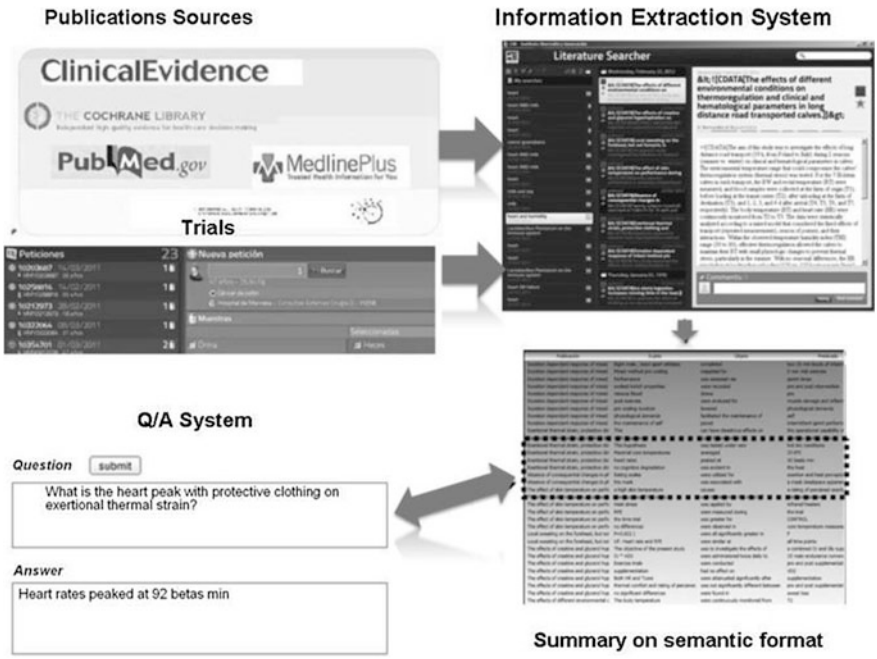

Fig. 3 Information retrieval system process (screen captures of the HENUFOOD developed platform)

6 Future Work

HENUFOOD is starting its last year of research and this is why although most of the procedures and functionalities previously described are fully functional prototypes, the intention is to use this year to increase the stability, performance and maturity of the whole system.

Another key point to improve is the better integration of all of its components and the integration of more external data sources in order to improve the central ontology. Furthermore, this ontology is continuously fed with new knowledge from the execution of clinical trials.

Finally, as the intention is let scientists use all the resources and take benefit of the knowledge base with questioning answering techniques using natural language processing with the avatar, empowering it with more flexibility and actions to take would be a good direction to follow.

Acknowledgments The present study is supported by the company Ibermática through HENUFOOD Project (CEN-20101016), part of the CENIT program from Spanish Minister of Economy and Competitiveness.

References

1. World Health Organization (2007) Nutrition health topics: background, the global burden of chronic. http://www.who.int/nutrition/topics/2_background/en/index.html. Accessed 28 Jan 2013
2. European Council and Parliament (2007) Guidance on the implementation of Regulation (EC) N°1924/2006
3. Demner-fushman D, Chapman WW, McDonald CJ (2010) What can natural language processing do for clinical decision support? J Biomed Inform 42:760–772
4. Liu K, Hogan WR, Crowley RS (2011) Natural language processing methods and systems for biomedical ontology learning. J Biomed Inform 44:163–179
5. Fayyad U, Grinstein GG, Wierse A (2002) Information visualization in data mining and knowledge discovery, Morgan Kaufmann, San Francisco, ISBN: 1-55860-689-0
6. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. In: American association for artificial intelligence FALL, pp 37–54
7. Suliman Z (2008) Knowledge discovery query language (KDQL). In:12th WSEAS international conference on computers
8. Alonso O, De La Peña P, Moreno A, Anzaldi y G, Domingo X (2012) Data mining and query answer techniques applied to a bio-nutritional trials focused expert system. In: Proceedings of the 15th international conference of the Catalan association for artificial intelligence (CCIA'12), pp 9–18
9. CONSORT (2010) CONSORT statement. http://www.consort-statement.org/. Accessed 01 Jan 2013
10. Borgelt C (2005) An implementation of the FP-growth algorithm. In: Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations OSDM, pp 1–5
11. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings international conference very large data bases, pp 487–499
12. CRISP-DM (2003) Cross industry standard process for data mining 1.0: step by step data mining guide. http://www.crisp-dm.org. Accessed 1 Jan 2010
13. U.S. National Library of Medicine (2012) SNOMED Clinical Terms® (SNOMED CT®), http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html. Accessed 29 Jan 2013

Springer

http://www.springer.com/978-3-319-02331-1

International Technology Robotics Applications
Proceedings of the 2nd INTERA Conference, held in Oviedo, Spain, March 2013
Gonzalez, I. (Ed.)

2014, XXII, 112 p. 76 illus., 53 illus. in color., Hardcover
ISBN: 978-3-319-02331-1