Automated Agricultural Robot and Sensor Data Collection and Analysis through a Biomass Feedstock Production Information System

Konstantinos Domdouzis[*]

College of Business, Technology and Engineering, Sheffield Hallam University, Sheffield, United Kingdom

Abstract: The increasing environmental pollution resulting from the use of non-renewable fossil fuels as well as the development of economic dependencies among countries because of the lack of such types of fuels underline the intense need for the use of sustainable forms of energy. Biomass derived biofuels provide such an alternative. The main tasks of biomass feedstock production are planting and cultivation, harvest, storage, and transportation. A number of complex decisions characterize each of these tasks. These decisions are related to the monitoring of crop health, the improvement of crop productivity using innovative technologies, and the examination of limitations in existing processes and technologies associated with biomass feedstock production. Other critical issues are the development of sustainable methods for the delivery of the biomass while maintaining product quality. There is the need for the development of an automated integrated research tool based on resilience and sustainability which will allow the coordination of different research fields but also perform research on its own. The specific tool should aim in the optimization of different parameters which specify the research done and in the case of biomass feedstock production; such parameters are the transportation of biomass from the field to the biorefinery, the equipment used, and the biomass storage conditions. This optimization would enhance decision making in the field of bioenergy production. Based on the need for such an automated integrated research tool, this paper presents an information system that provides automated functionalities for better decision making in the bioenergy production field based on the collection and analysis of agricultural robot and sensor data.

Keywords: Bioenergy, Agricultural Robot, Sensor, Information System, Automation, Integration, Databases, Data Mining, Simulation.

1. INTRODUCTION

There is the need for a systems informatics infrastructure which will be capable of providing appropriate management of the existing and the generated knowledge. The ultimate target of systems informatics is the improvement of the decision-making process.

The long-term goal of the presented research is the provision of concurrency in the collection and integration of information from the research and development of different tasks in the bioenergy supply chain research and development. For this purpose, it is required that software engineering techniques which will perform problem requirements analysis and identification of the systems informatics architecture, are used. The central hypothesis of the presented research is that an information system with integrated functionalities can improve the process of data collection. Such an infrastructure can also provide more automation in the development of computational models and also enable the real-time collection of information. The rationale is that the provision of automation through the principles of object-oriented analysis, the identification of relationships between the attributes of these objects, the concurrent integration of data, the use of advanced software technologies, such as databases and visualization tools, and the combination of data for better decision support will improve life-cycle and gap analysis, the estimation of energy and mass balance, as well as the peer review of information.

The overall aim of the research is the development of an information management infrastructure which will aid in decision making during the research and development of biomass feedstock production systems, provide improved optimization processes and effectively handle data between the biomass feedstock production tasks and subtasks with focus on the acquisition of data from advanced technologies such as agricultural robots and sensors.

2. BIOFUELS AND THEIR SIGNIFICANCE

The energy needs of modern societies depend on fossil fuels. An example is the reliance of the United States on fossil fuels as it satisfies 81% of its energy needs from them [1]. In 2019, the imported crude oil reached about 74% of the U.S. total gross petroleum imports while the non-crude oil was 26%. [2]. The use of oil in transportation is approximately 26% of all energy use [3]. This dependence however has resulted to the exhaustion of oil resources and the creation of global environmental problems.

*Address correspondence to this author at the College of Business, Technology and Engineering, Sheffield Hallam University, Sheffield, United Kingdom; E-mal: k.domdouzis@shu.ac.uk

The resulting dramatic increase in oil prices has led to the increase in the use of biofuels. Biofuels are liquid, solid, or gaseous fuels produced by organic matter. Global biofuel production is expected to increase again by 2026 since it fell in 2020 due to the Covid-19 pandemic [4]. The majority of the biofuel production is done in the United States, Europe and Brazil [5]. Currently, biofuels fall into three main generations as outlined below:

- First Generation: This category includes petroleum gasoline and petroleum diesel substitutes. First generation biofuels are ethanol and butanol. Ethanol is produced by sugar fermentation extracted from sugar cane or sugar beets, or sugar extracted from starch contained in maize kernel or other crops which contain starch [6]. Similar processing can produce butanol.

- Second Generation: In this category, biochemically produced petroleum-gasoline, thermo-chemically produced and thermo-chemically produced petroleum-diesel substitutes are included [6].

- Third Generation: Algae biofuels are included in this category [6]. Microalgae are capable of producing 15-300 times more oil for biodiesel production than traditional crops [7]. Microalgae have a very short harvesting cycle, and this allows multiple or continuous harvests with increased yields [8].

The use of perennial grasses for the production of bioenergy is a promising renewable energy option. Perennial grasses can be used for the production of electricity while the reduction of the use of fossil fuels will result to ecological benefits. Examples of perennial grasses are miscanthus and switchgrass. Switchgrass can control soil erosion and with its extensive network of roots, it can reduce raindrop impact and erosion [9]. Miscanthus can be used for the generation of electricity and the direct heating of homes and businesses [10]. It is a low-maintenance, perennial grass which is characterized by long-life expectancy. Nitrogen and other nutrients are moved into the rhizomes during the growing season for next year's growth.

3. BIOMASS FEEDSTOCK PRODUCTION

Biomass feedstocks are used in the production of liquid transportation fuels. The biomass feedstock production system includes a number of inputs and factors for the conversion of biomass into energy.

The different phases of biomass feedstock production are presented below:

- Pre-Harvesting: This stage is the initial stage of biomass feedstock production and includes operations such as crop selection, soil preparation, and planting. Since these operations are very significant for the realization of the next stages in biomass feedstock production, advanced remote sensing and precision agriculture technologies are employed.

- Harvesting: Harvesting is one of the major phases of biomass feedstock production. Harvesting can include mowing, raking, baling, staging, and loading. The type of equipment defines the time and cost of harvesting.

- Transportation: This task involves the transportation of biomass to the biorefinery or to a centralized facility for storage purposes. There are different ways of transportation of the biomass and these include pipelines, rail, or trucks.

- Storage: This task includes on-farm or centralized storage. Different parameters, such as microbial activity, dry matter loss, temperature, and moisture content, need to be considered when the storage occurs in different time periods.

4. AGRICULTURAL SYSTEMS INFORMATICS

Systems Informatics is the combination of social and organizational informatics. Social informatics involves the design, implementation, and use of Information and Communication technologies over a wide range of social settings [11]. Organizational informatics refer to the use of social informatics techniques within the limits of an organization [12].

An example of the use of systems informatics is the development of information systems. The main activities of information systems are data acquisition and data handling. Data acquisition is achieved through data suppliers, such as research organizations and automatic monitoring stations. Data handling is realized in computational centers while the use of information is done by decision makers.

This section provides examples of how information systems and associated methodologies can be used in different applications in agriculture. The provided examples attempt to show how information systems correspond to the main principles of systems informatics.

4.1. Automation-Culture-Environment oriented Systems (ACESYS)

The concept of Automation-Culture-Environment oriented systems (ACESYS) facilitates abstraction of controlled environment plant production systems, describes the relationships among sub-systems or system elements, identifies information and protocols required for system analysis, and develops data processing algorithms. Automation refers to the data processing and task execution related to the operation of the system. Culture defines the number of factors that can describe the biological growth of the plants. Environment refers to the environment of the plants. This environment is characterized by different types of conditions, such as climatic and structural conditions. System refers to the definition of the system and produces conclusions related to the system's productivity and reliability. The ACESYS concept provides a modular description of the system by defining a number of classes and objects related to the system. Specifically, an object-oriented approach based on ACESYS enables the development of a number of foundation classes that can be used for the description of the different elements of closed plant production systems [13].

4.2. The Tropos Methodology

The Tropos methodology is an agent-based software development methodology. It is based on the use of knowledge level concepts, such as actors, goals and plans and the dependency between them, and also on the critical role which is implemented to the preliminary phase of requirements analysis. The Tropos Methodology includes five software development phases which are the identification of early and late requirements, the design of the system's architecture, the design of the agent implementation architecture, and system implementation. The design of the system's architecture is the most significant phase since it allows the definition of the parts of the system which are represented as actors. Each actor is characterized by a set of capabilities and a set of social capabilities which allows the coordination with other actors [14]. "The Tropos methodology allows the extraction of requirements and this is very useful in the development of agent-based systems in Agriculture".

4.3. Agriculture 4.0

Agriculture 4.0 is considered the coming agricultural revolution which will be green, and it will use science and technology as its core elements. In Agriculture 4.0, there will be no longer dependence on the use of water, fertilizers or pesticides. Instead, new technologies will be integrated to the agricultural supply chain. Examples of such technologies are hydrophonics seawater technologies, bioplastics, drone technologies, data analytics, 3D Printing, and Internet-of-Things (IOTs). Examples of the use of such technologies in modern farming are data-driven farming through the analysis of agricultural information (eg. types of seeds, weather data, market prices), AI-powered chatbots that could provide recommendations to farmers about specific problems, the production of 3D maps for early soil analysis and the use of drones for the collection of data related to irrigation and nitrogen levels, and the use of Blockchain for the identification of problems in the agricultural supply chain that lead to waste or for the identification of any food fraud. Furthermore, crop spraying and monitoring can be improved through the use of drones that will scan the ground in real-time while time-series analysis can show the gradual development of a crop [15].

4.4. Decision Support System for Water Resource Management

Water resources are under increasing pressure and this is because of the increment of economic activities, population growth and climate change. Water resource management includes two components that are related to each other. The first component is the management of the sources of the resource and the second component is water demand. In some basins, water management is focused on the satisfaction of agricultural demands. This is for example the case of the Barbate River Basin in Cádiz, Spain. The efficient use of water in agriculture has been studied extensively. The optimization of water can be analyzed based on its availability. In order for the Barbate River Basin to be modelled, two decision-support systems (DSSs) were used: the AQUATOOL and the SIGMES. AQUATOOL was developed by the University of Valencia, Spain and it was originally designed for the planning stage of decision-making. SIGMES uses an objective function in which each term is an objective function that corresponds to each element of the system. A simulation model was built using the SIGMES module of AQUATOOL and this model shows the complex interactions among all the system's

elements. Also, different simulation scenarios were generated for the optimization of water resources in the Barbate River Basin. Based on the use of DSSs, specific characteristics of the basin were analyzed and aspects that need improvement were identified. Also, appropriate management strategies which are characterized by the lowest cost of implementation and the fewest environmental impact were specified [16].

5. BIOMASS FEEDSTOCK PRODUCTION ENGINEERING INFORMATION SYSTEM (BFPE-IS)

The Biomass Feedstock Production Engineering Information System (BFPE-IS) includes three layers: the concept diagrams which show the different tasks associated with bioenergy production scenarios and their respective sub-tasks, the Engineering Solutions for Biomass Feedstock Production Engineering Systems Informatics Architecture which shows the software technologies that consist of the skeleton of the presented information system, and the Biomass Feedstock Production Engineering – Application Programming Interface (BFPE-API) that enables the processing of information for data mining purposes. The different elements of the BFPE-IS are presented in detail in the next sections.

5.1. Concept Modeling

Concept mapping can be used to improve collaboration and communication in a project team by visually capturing the relationship between scenarios and features. A concept mapping approach has been used in order to validate its effectiveness. The development of the product was done over four one-week iterations using Java as the programming language, MySQL as the database, Eclipse as the development environment, and CMapTools to manage the concept mapping activities. Participants included four members with agile usability training who acted as the project team in different roles. A local non-profit mediation agency served as the client and end user. There was provision of weekly surveys to participants in order to capture each participant's perspective on the project and the concept mapping approach [17]. The participants found the visual nature of concept mapping helpful in enhancing collaboration among them.

Fuzzy-logic Cognitive Mapping is used in order to describe differences in mental models among stakeholders in a conservation agriculture development project. This type of modelling identifies representations of mental models of factors and the relationships between environmental conditions (eg. soil moisture), farm-based dynamics (eg. crop yield) and different technologies. In order to develop the mental model representations, a number of interviews with farmers were conducted so that the different factors that affect farm dynamics to be identified. Based on these interviews, a quantitative survey was used in order to develop a cognitive map that would be used by experts and individuals [18].

Concept diagrams have been used for the representation of the different tasks associated to biomass feedstock production. The design of the diagrams is based on a top-to-down hierarchy in which the lower levels represent the sub-tasks of a specific task and the lowest levels correspond to the technologies used for the realization of the specific tasks and sub-tasks. Each box includes a number of attributes related to the specific task, sub-task, and technology. The identification of the different attributes is important as it helps in the development of connections among the tasks and sub-tasks.

The development of the concept diagrams was based on the collection of information from the literature and interviews with expects in bioenergy production. The concept diagrams are dynamic and this means that the information included in them can be changed any time depending on new information collected by the literature. The data collected from the Concept Maps are in addition to the data collected by agricultural robots and sensors.

Furthermore, through the development of concept diagrams, different researchers from different stages of Biomass Feedstock Production can provide a variety of information about the tasks, sub-tasks, technologies and their associated attributes used in Biomass Feedstock Production.

5.2. Biomass Feedstock Production Engineering Systems Informatics Architecture

The core element of the systems informatics architecture is a number of MySQL/Microsoft SQL Server or Oracle databases called Biomass Feedstock Production Databases (BFPDs). The structure of the databases is based on the concept diagrams designed for each task of the biomass feedstock production. Each of the databases can also refer to a specific sub-task. The concept diagrams provide the data-input protocol or in other words, the order in which the data are inserted in the database and also the hierarchical connections which characterize the structure of the

database. The BFPDs can collect data automatically by different types of sensors such as soil moisture sensors or pH sensors. These sensors form wireless sensor networks that can communicate with the BFPDs. Also, agricultural robots can send data to the BFPDs through the use of Raspberry PIs and Arduinos.

A set of graphical user interfaces have also been developed in order to allow the dynamic manipulation of the collected data that are stored in the database(s) and also the initiation of data discovery. The interfaces can additionally initiate the connection to a number of simulation and optimization software packages, such as MATLAB and GAMS. These software packages are used for the development of models which accept data by the database. These models also provide data back to the database. The interfaces along with all the functionalities they offer consist of an Application Programming Interface (API) used for Biomass Feedstock Production data analysis. This API is referred as Biomass Feedstock Production Engineering (BFPE)-Application Programming Interface (API).

The ultimate purpose of systems informatics is the provision of automated search and analysis tools. The existing and the generated data will allow the peer review of information and the realization of different types of analyses, such as life-cycle and gap analysis. Extensibility is also a major target as it will allow the addition of new users who can enable the generation of more information and the expansion of the limits of the current research.

5.3. Description of the BFPE-Application Programming Interface (API)

The Biomass Feedstock Production Engineering (BFPE)-Application Programming Interface (API) is a way for users to communicate with the Biomass Feedstock Production database(s) (BFPD). The functionality of the API can be classified in three layers. The most basic layer is the database data access which allows the user to connect to the BFPD(s) using a number of different programming languages. The use of different programming languages allows users of different computing backgrounds to access the database. The next layer is the database data handling. This layer includes operations such as insertion of data in the database, deletion of data from it, and updating of its data. The database data discovery layer is the top layer of functionality of the API and includes advanced data mining and statistical analysis operations. Additionally, visualization and optimization techniques are part of this layer.

The BFPE-API is written in different programming languages and it is dynamic. The programming languages used for the development of the API are Java, Visual Basic, Python, PHP, Perl, Visual C#, C++, Ruby-on-Rails. A number of software tools are used as part of the API and these are MATLAB, Java APIs such as Lucene, JFreeChart, and the Generic Algebraic Modeling System (GAMS). The BFPE-API and its interaction with the MySQL database(s) is shown in Figure **1**.

6.   EXAMPLES OF BFPE PLATFORM FUNCTIONALITIES THAT LEAD TO DATA MINING

This section presents specific examples of the API functionalities that enable the efficient processing of biomass feedstock processing information. These functionalities are presented in detail below.

6. 1. Dynamic Data Formulation Through Concept Mapping

Concept Maps are the points of collection of data from users that work on different aspects of biomass feedstock production engineering. The Concept Maps are also the basis for the automated development of a dynamic Entity-Relationship (ER) Diagram which in turn is the basis for the development of the relational databases used (as part of the BFPE system) for data storage and processing. The use of Visual Paradigm Enterprise Edition allows the development of ER Diagrams connected to the database(s). In this case, the users can develop the ER model on Visual Paradigm Enterprise Edition, initiate the connection to the database and update its tables dynamically through Visual Paradigm. This is an example of data-driven automation can be used for further robotic process automation.

6.2. Data Association & Classification

Data Association refers to the grouping of related data elements. Data Classification is the process of organizing data in different categories so that their analysis is realized more efficiently. Data Association and Classification are achieved through specific algorithms or the use of libraries that execute automatically these algorithms. In the case of the BFPE-API, this is achieved through the use of PHP and Python libraries. For example, Scikit-Learn is a Python library that allows data classification through the use of a number of algorithms such as decision trees or logistic regression. A variety of data that are stored in BFPD databases have been associated or classified based on their attributes. In this case, data are
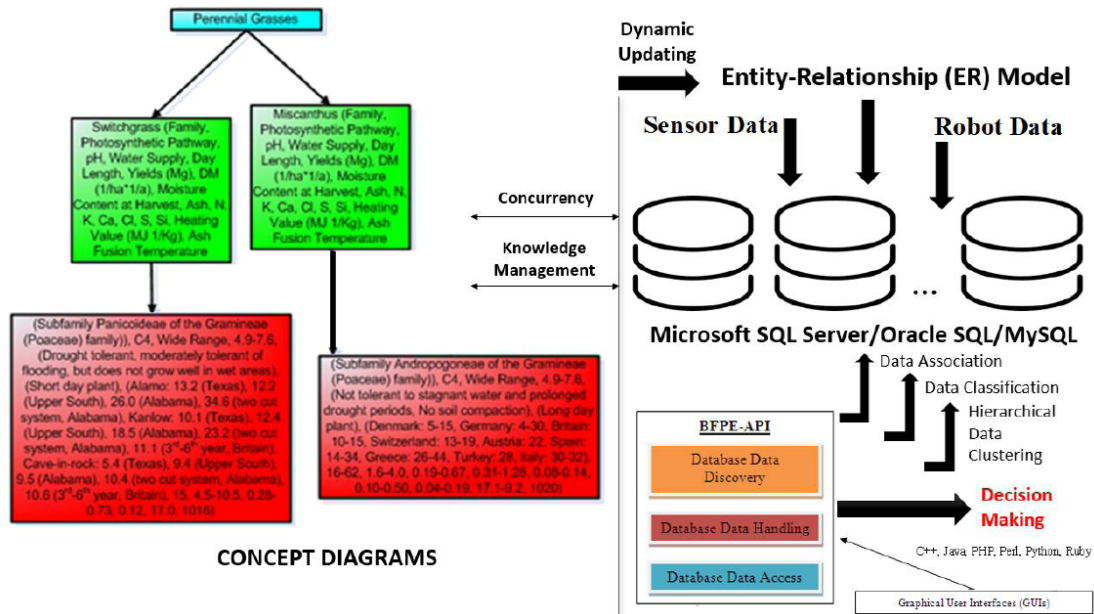
Figure 1: Systems Informatics Architect [Adapted from [5]].

clustered based on the fact they present similarities to one or more attributes.

6.3. Hierarchical Data Clustering

There are two types of hierarchical clustering: the agglomerative and the divisive. The presented information system uses the agglomerative hierarchical clustering. The agglomerative hierarchical clustering is based on the merging of the closest pair in the set of the current clusters into a new cluster. Two clusters are merged when they present the largest similarity, thus the smallest distance. The distance matrix which has been created and which shows the distances among the initial clusters, needs to be updated. There are different methods for measuring the inter-cluster distance. For example, the distance between two clusters can be measured by calculating the distance between the closest pair of data objects that belong to different clusters. Other methods that are used for the measurement of the inter-cluster distance are the calculation of the farthest pair of data objects, the calculation of the average distance of all pairs of data objects belonging to different clusters and the calculation of the distance of the centroids of the two clusters. A distance matrix using the distance values needs to be constructed. When two clusters are merged, they are removed from the matrix and the new cluster is placed instead. The distances of the new cluster from all the other clusters have to be evaluated and the distance matrix needs to be updated again. This procedure is repeated till one single cluster is formed [19]. The Python Scikit-Learn library has been

used for the realization of hierarchical data clustering.

7. CONCLUSIONS

The examples of the BFPE-API functionalities which are presented in this paper, show how the analysis of agricultural robot and sensor data can be automated through the integration of database technologies with specific software tools. The presented information system enables the concurrent exchange of data among various users with respect to their different information technology backgrounds. Specifically, the use of different programming tools and languages for the development of the information system allows users to communicate with each other automatically. This is achieved through the use of Graphical-User Interfaces (GUIs) in which the users can insert, check, update and delete data from the BFP databases. In this case, the BFPD databases act as a centralized point of data collection and analysis.

The BFPE-API functionalities, which are presented in this paper, show how automated data analysis and search can be achieved. These functionalities allow the dynamic updating of the structure and the relationships of different types of information and the identification of associations among them. The identification of these associations allows the development of new scenarios on bioenergy production simulation. The same also applies for the classification of the data that are included in the BFP databases. Specifically, both association and classification have been proven very useful when large amounts of data are included in the

databases as they can allow the identification of hidden relationships within data based on a single or more attributes or the clustering of data again based on certain attributes.

Furthermore, the BFPE-API functionalities are good examples of knowledge management and concurrent engineering, since they allow the bi-directional exchange/modification of data among users. Software engineering is also achieved in a more automated way. For example, the use of concept diagrams in combination with the capabilities offered by Visual Paradigm allows the updating of the data of the database(s) in an automated way.

ACKNOWLEDGMENT

REFERENCES

[1]   The National Academies of Sciences, Engineering, Medicine (2021) Our Energy Sources - Fossil Fuels. Available at: http://needtoknow.nas.edu/energy/energy-sources/fossil-fuels/ (Accessed: 23 August 2021)

[2]   U.S. Energy Information Administration (2021) How much petroleum does the United States import and export? Available at: https://www.eia.gov/tools/faqs/faq.php?id=727&t=6 (Accessed: 23 August 2021)

[3]   U.S. Energy Information Administration (2021) Use of energy explained - Energy use for transportation. Available at: https://www.eia.gov/energyexplained/use-of-energy/transportation.php (Accessed: 23 August 2021)

[4]   Argus (2021) World biofuels output up to 3.3mn b/d by 2026: IEA. Available at: https://www.argusmedia.com/en/news/2196777-world-biofuels-output-up-to-33mn-bd-by-2026-iea (Accessed: 22 August 2021)

[5]   International Energy Agency (IEA) (2021) Transport Biofuels. Available at: https://www.iea.org/reports/transport-biofuels (Accessed: 22 August 2021)

[6]   RA. Lee and JM. Lavoie. 'From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity', Animal Frontiers, 2013; 3(2): pp. 6-11
https://doi.org/10.2527/af.2013-0010

[7]   AEM. Hussian (2018) 'The Role of Microalgae in Renewable Energy Production: Challenges and Opportunities', in Turkoglu M. (ed.) Marine Ecology - Biotic and Abiotic Interactions. Intech Open
https://doi.org/10.5772/intechopen.73573

[8]   Dragone, G., Fernandes, B., Vicente, A.A. and Teixeira, J.A. (2010) 'Third generation biofuels from microalgae', Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, pp. 1355-1366

[9]   Department of Plant and Soil Sciences, Oklahoma State University (2021) Environmental Impacts. Available at: http://switchgrass.okstate.edu/environmental-impacts/environmental-impacts (Accessed: 22 August 2021)

[10]   Kinhal V. (2021) Safe and sound: Decentralisation with Miscanthus giganteus. AgriKinetics. Available at: https://www.agrikinetics.com/miscanthus-giganteus/decentralisation-with-miscanthus/ (Accessed: 22 August 2021)

[11]   IGIGlobal (2021) 'What is Social Informatics'. Available at: https://www.igi-global.com/dictionary/framework-analyzing-social-interaction-using/27363 (Accessed: 05 July 2021)

[12]   Meyer ET, Shankar K, Willis M, Sharma S, Sawyer S. (2019) 'The social informatics of knowledge', J Assoc Inf Sci Technol. 2019 Apr; 70(4): 307-312
https://doi.org/10.1002/asi.24205

[13]   Kozai T, Fujiwara K, Runkle ES. (Eds) LED Lighting for Urban Agriculture. Springer Singapore, Singapore 2016.
https://doi.org/10.1007/978-981-10-1848-0

[14]   Tropos (2021) The Tropos Methodology. Available at: http://www.troposproject.eu/node/93 (Accessed: 10 May 2021)

[15]   De Clercq M, Vats A, Biel A. (2018) 'AGRICULTURE 4.0: THE FUTURE OF FARMING TECHNOLOGY'. World Government Summit. Available at: https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2021/apr/agriculture-4-0-the-future-of-farming-technology.pdf (Accessed: 12 May 2021)

[16]   Ruiz-Ortiz V, García-López S, Solera A, Paredes J. 'Contribution of decision support systems to water management improvement in basins with high evaporation in Mediterranean climates'. Hydrology Research 2019; 1, 50(4): 1020-1036
https://doi.org/10.2166/nh.2019.014

[17]   Barksdale, J. and D. Scott McCrickard (2010) 'Concept Mapping in Agile Usability: A Case Study', CHI 2010: Usability Methods and New Domains, April 10-15, 2010, Atlanta, GA, United States, pp. 4691-4694
https://doi.org/10.1145/1753846.1754212

[18]   Gray SA, Zanre E and SRJ Gray. (2014) 'Fuzzy cognitive maps as representations of mental models and group beliefs: theoretical and technical issues'. pp. 29-48 in E. I. Papageorgiou (Ed.). Fuzzy cognitive maps for applied sciences and engineering -from fundamentals to extensions and learning algorithms. Springer, Heidelberg, Germany.
https://doi.org/10.1007/978-3-642-39739-4_2

[19]   Sharma, P. (2019) 'A Beginner's Guide to Hierarchical Clustering and how to Perform it in Python', AnalyticsVidhya. Available at: https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/ (Accessed: 15 July 2021)