



Global Journal on Technology



Issue 7 (2015) 07-13

Selected Paper of 4th World Conference on Innovation and Computer Sciences (INSODE-2014)
11-12 April 2014, Quality Hotel Rouge et Noir Conference Center, Roma, Italy

Integrating expert information on tobacco production

Kolyo Onkov*, Department of Mathematics, Computer Science and Physics, Agricultural University, Plovdiv, Bulgaria.

Hristo Bozukov, Tobacco and Tobacco Products Institute, Agricultural Academy, Plovdiv, Bulgaria.

Maria Kasheva, Tobacco and Tobacco Products Institute, Agricultural Academy, Plovdiv, Bulgaria.

Suggested Citation:

Onkov, K., Bozukov, H. & Kasheva, M. Integrating expert information on tobacco production, *Global Journal on Technology* [Online]. 2015, 07, pp 07-13. Available from: www.awer-center.org/pitcs

Received December 10, 2013; revised February 16, 2014; accepted March 17, 2014.

Selection and peer review under responsibility of Assoc. Prof. Dr. Özcan Asilkan, Akdeniz University, Turkey.

©2015 SPROC LTD. Academic World Education & Research Center. All rights reserved.

Abstract

Information and knowledge on tobacco production is multi-aspect, heterogeneous, semi-structured and stored in different information sources. The paper presents a multidimensional database where data cubes are used for database building, thus giving the opportunity to store contextually related information on tobacco production in hierarchical structures. The relational database model presents four dimensions: Phytopharmacy, Varieties, Statistics and Legislation. The concept is used for integration of expert information on tobacco production in Bulgaria regarding: pests and permitted for use pesticides; tobacco groups and varieties including description of their biological features, chemical composition of tobacco and tobacco smoke, etc.; statistical time series on areas and quantity production; legislative documents. The presented solution is: a) hierarchically structured multidimensional database which integrates extracted information on tobacco production from different sources; b) open system as it regards to the number of dimensions and updates of stored information; c) platform for building databases which can integrate information on diverse crops in other countries, as well; d) user-friendly environment which helps the users to avoid the difficulties associated with finding and working with information in different files and formats. This solution increases expertise of scientists, teachers, experts and managers in their academic work and in the process of decision-making.

Keywords: multidimensional database, data cubes, tobacco varieties, pest management, statistical time series, user friendly environment.

* ADDRESS FOR CORRESPONDENCE: **Kolyo Onkov**, Agricultural University - Department of Mathematics, Computer Science and Physics, 12 Mendeleev, 4000 Plovdiv, Bulgaria. *E-mail address:* kolonk@au-plovdiv.bg / Tel.: +359-32-654472

1. Introduction

Information integration is the problem of combining datasets residing at disparate sources, and providing the user with unified view of these data [1, 2]. Technologies usually used for information integration are sets of relations in multidimensional databases, metadata techniques in data warehouses and interconnections in Web environments [3, 4, 5]. During the last decade Geographic Information Systems and Cloud Computing Information as a Services are considered as typical systems where large scale datasets of graphical, numerical and text data are integrated [6, 7, 8].

Expert information and knowledge on tobacco production is stored in different information sources. Its integration is needed for academic, administrative and management purposes. The paper aims to present hierarchically structured multidimensional database, thus giving the opportunity to integrate related expert information on tobacco production, including the application on tobacco in Bulgaria. The paper also discusses the opportunity this solution to be used as a platform for building databases which can integrate information on diverse crops in other countries, as well.

2. Information sources and database modeling

Expert information on tobacco production is heterogeneous, spatial and temporal, usually found in semi-structured form at different pre-existing information sources: publications by scientists and experts in universities and research centres, governmental institutions – Ministry of agriculture, statistical service institutions, etc. Expert information on tobacco is multi-aspect: biological, chemical, physical, statistical, technological and legal. This information does not changes often. Consequently, it is not needed to create mediated schema with virtual relations on which users will pose queries [3]. A more proper solution is the open relational schema of database regarding the dimensions, attributes and updates. Besides, the database designer should accept the end-users' attitudes and requirements regarding substantial relationships between entities and information visualization.

Following these thoughts the multidimensional database for storing expert information on tobacco production is designed through using data cubes. This approach gives the opportunity to store contextually related information in hierarchical structures. Multidimensional databases consist of one or collection of data cubes [4]. Figure 1 shows ERA (entity relationship attributes) model of multidimensional database for storing expert information on tobacco production.

The relational database model presents four dimensions: Phytopharmacy, Varieties, Statistics and Legislation, organized in three data cubes. The data regarding dimensions Varieties and Statistics are related in one cube because of sharing identical attributes – VarietyGroups and VarietyList. ERA database models operate with two types of tables: dimension and fact tables. Dimension tables store attributes and relationships among tables. Fact tables are placed on the lowest level of hierarchy and contain expert information on tobacco production and foreign keys which refer to primary keys in the dimension tables. The hierarchical organization of multidimensional database has the following features: a) It is a systematic solution with high operability that facilitates flexible data access, visualization, querying and updating; b) It abstracts information classification, substantial relations among entities as well as spatial and temporal details in datasets.

Expert information on tobacco production is stored in cubes as follows:

- **Phytopharmacy.** Expert information on tobacco pest management can be generally found in databases supported by international or governmental institutions (EU Pesticides DB, PAN pesticides DB, etc.), in the sites of big pharmaceutical companies (BASF, Syngenta, Bayer CropScience, etc.) and in Crop Protection Compendium of CABI. Let's note that each country has rights to impose own rules, regulations and limitations on pest management. Consequently, authorized institution of the country publishes guides and reference books on permitted pesticides for use. This information, usually in

semi-structured form, is detailed and widely used in practice. All mentioned above concern the expert information on tobacco pesticides use. The expert character of this information presumes contextually related words and phrases (key words) – the names of tobacco pest groups and pests and pesticides (figure 1). For the developed multidimensional database it is used: a) biological classification of pests on tobacco in 6 groups: Fungi&Viruses\$Bacteria, Insects, Weeds, Nematodes, Rodents and Pheromones; b) coding of key words for building relationships following the hierarchy of classification and substantial relations „pest group – pest – pesticide” [9, 10]. The names of pest groups and pesticides are stored in dimension tables of the database and used as primary keys, while the quantitative and qualitative characteristics of each pesticide: active ingredient, dose and minimum lethal dose, PHI days, use category and application are attributes of the fact tables. It is important to point out that this approach provides end users with the opportunity to work with real terms of phytopharmacy: names of tobacco pests and pesticides, but not with their codes.

•**Varieties and Statistics.** The three tobacco variety groups are taken into consideration: Flue-cured, Light air and Oriental. Then the list of varieties corresponds to the specific tobacco varieties in a given country. The varieties are described through classified expert texts presenting information and knowledge on: biological features of tobacco crop (genealogical, morphological and botanical), soil and climate requirements, conditions for production (plant population, breaking, picking, yield, curing and as well as classification and packing), chemical composition of tobacco and smoke (nicotine, soluble carbohydrates, total nitrogen and tar). As it concerns statistics, for each variety time series on tobacco production are stored by harvested areas and quantities. Time series are placed in fact tables vertically. Therefore, the updating requires adding only one record in each fact table of the database.

•**Legislation.** This part of database is designed for storing two types of text materials regarding tobacco production: laws and normative documents. The texts are saved in external files which are accessible through database techniques and wide spread software system (Word, Adobe Acrobat etc.). PDF files are preferable to guarantee reading the legislative text materials on tobacco in original form.

Finally, it is necessary to point out that the proposed concept for multidimensional database operates with different types of information: text, numerical and graphical.

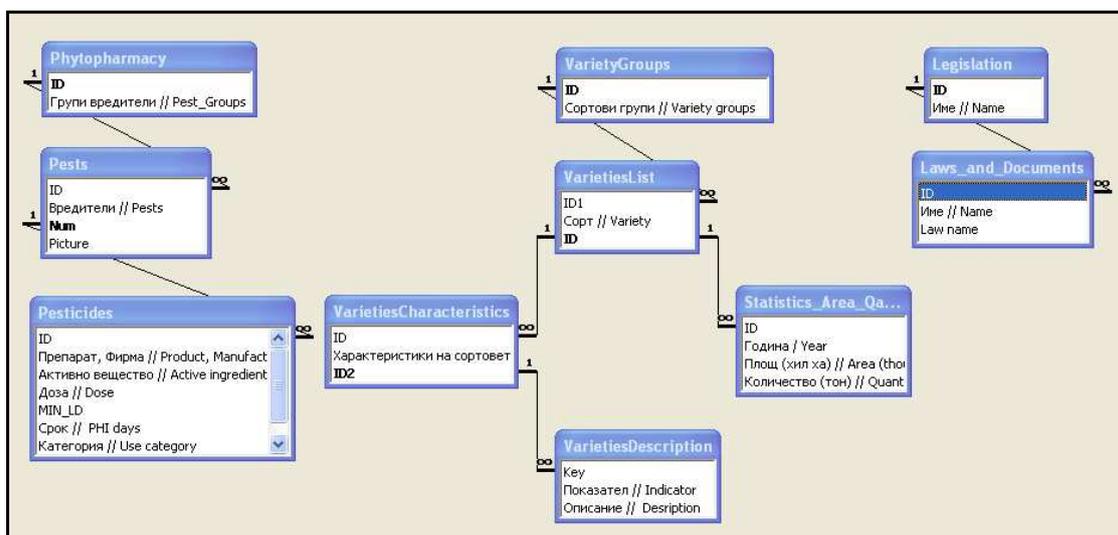


Figure 1. ERA model of multidimensional database

3. Integration of expert information on tobacco production in Bulgaria

The concept for integration of expert information on tobacco production presented in the previous chapter is applied for creation of Tobacco_BG database. For this purpose biological, chemical, physical, statistical and legal information is used. It is presented in diverse file formats and stored in different information sources.

Annually Ministry of Agriculture in Bulgaria prepares a reference book including list of permitted pesticides for crops protection in the country [11]. The book content is based on state laws and regulations, as well as it reflects the knowledge of the national experts. For building the cube Phytopharmacy of the Tobacco_BG database the data from this reference book and expert attitudes regarding tobacco pests and pesticides are integrated [12]. Figure 2 illustrates data access and visualization through relationships in the sequence „pest group – pest – pesticide”. After the choice of pest “Peronospora tabacina” from the group (Fungi & Viruses\$ Bacteria), its image and list of all corresponding pesticides are extracted from the database. The quantitative and qualitative characteristics of each pesticide – active ingredient, dose, minimum lethal dose (MIN_LD), PHI days, category and instructions for the pesticides application are presented. Therefore, database techniques provide extraction of comprehensive and useful information for the tobacco crop protection specialist.

Групи вредители // Pest_Groups							
- Патогени // Fungi&Viruses\$Bacteria							
Вредители // Pests				Picture			
- Мана // Peronospora tabacina							
Препарат_Фирма // Product_Manufacture	Активно вещество // Active ingredient	Доза // Dose	MIN_LD	Срок // PHI days	Категория	Приложение // Application	
АКРОБАТ МЦ БАСФ Агро Б.В. // Acrobat MZ BASF Agro B.V.	90 г/кг диметоморф + 600 г/кг манкоцеб // 90g/kg dimethomorph	0.2%	1750	14	3	При поява на първи петна. ИТ 10-14 дни. РР 40.	
АКРОБАТ МЦ БАСФ Агро Б.В. // Acrobat MZ BASF Agro B.V.	60 г/кг диметоморф+400 г/кг меден оксихлорид // 60g/kg	0.25% (250 г/дка)	2000	20	3	При поява на първи петна. ИТ 10-14 дни. РР 40.	
АКРОБАТ ПЛЮС БАСФ Агро Б.В. // Acrobat PLUS BASF Agro B.V.	90 г/кг диметоморф + 600 г/кг манкоцеб // 90g/kg dimethomorph	200 г/дка	1750	14	2	При поява на първи петна. ИТ 10-14 дни. РР 40.	
АНТРАКОЛ 70 ВП, Байер КронСайанс // ANTRACOL 70WG, Bayer CropScience	700 г/кг пропинеб // 700 g/kg propineb	0.20%	5000	7	1	При поява на първи петна. ИТ 7-10 дни.	
ВЕРИТА ВГ Байер КронСайанс // VERITA WG Bayer CropScience	44 г/кг фенамидон + 667 г/кг алуминиев фосетил	0.15%	2028	30	3	При поява на първи петна. ИТ 7-10 дни.	
ДИТАН М-45 ВП Индофил Индъстриъз Лимтед // DITAN M-45 WP Indofil	800 г/кг манкоцеб // 800 g/kg mancozeb	0.2 %	5000	20	3	При поява на първи петна. ИТ 7-10 дни. РР 30-60л/дка.	
КОРСЕЙТ М ДФ Дюпон // CURZATE M DF DuPont IO	4% симоксанил + 40% манкоцеб // 4% сутохапил + 40% mancozeb	0.3%	2000	20	3	При поява на първи петна. ИТ 7-10 дни. РР 30.	
ПОЛИРАМ ДФ БАСФ // POLIRAM DF BASF	70% метирам // 70% metiram	0.2%	5000	14	3	При поява на първи петна. ИТ 7-10 дни. РР 40.	
РИДОМИЛ ГОЛД МЦ 68ВГ Синджента // RIDOMIL GOLD MZ 68WG Syngenta	4% мифаноксам + 64% манкоцеб	250 г/дка	2000	14	3	1-во, при поява на първи петна. ИТ 10-14 дни. РР 40.	
*							
+ Брашнеста мана // Powdery mildew							
+ Кафяви листни петна // Alternaria spp.							

Figure 2. Tobacco_BG database: Phytopharmacy cube

Information needed for building the Varieties cube has been collected by biologists and chemists from Agricultural Academy, Tobacco and Tobacco Products Institute. The three varieties groups mentioned in chapter 2 are taken into account – Flue-cured, Light air and Oriental (figure 3). Expert information on 13 original Bulgarian tobacco varieties from the Oriental group is stored in the database. The names of these varieties correspond to the specific geographic regions where they are cultivated. For each variety, information on biological and chemical characteristics of tobacco crop, soil and climate requirements and conditions for production, is stored. Figure 3(a) presents detailed information on chemical composition of tobacco and smoke of Oriental variety “Melnik”. The statistical data on tobacco production is stored in the database in the form of spatial time series. Hierarchical structures of Tobacco_BG database enable the end users to view more than one relational object simultaneously. Figure 3(b) illustrates annual time series data on harvested area and production quantity for two Oriental varieties: “Melnik” and “Ustina”.

Сортови групи // Variety groups		Сортови групи // Variety groups	
+ Флу кюрд // Flue-cured		+ Флу кюрд // Flue-cured	
+ Лайт еар // Light air		+ Лайт еар // Light air	
- Ориенталски - Басма // Oriental - Basma		- Ориенталски - Басма // Oriental - Basma	
Сорт // Variety		Сорт // Variety	
+ Джебел / Djebel		+ Джебел / Djebel	
+ Неврокоп // Nevrokop		+ Неврокоп // Nevrokop	
+ Дупница // Dupnitsa		+ Дупница // Dupnitsa	
▶ - Мелник // Melnik		▶ - Мелник // Melnik	
Характеристики на сортовете // Varieties characteristics		Година / Year Площ (хил ха) // Area Количество (тон) // Quantity	
+ Специални характеристики / Special characteristics		▶ 1998 1.61 1880	
+ Условия на производство / Conditions for production		1999 0.87 836	
▶ - Химичен състав на тютюна и тютюневия дим / Chemical composition		2000 0.94 844	
Показател // Indicator Описание // Description		2001 0.86 1139	
Никотин (тютюн) // Nicotine (tobacco)	0,9 - 1,6%	2002 0,92 970	
Разтворими въглехидрати // Soluble carbohydrates	9 - 12,2%	* 0 0 0	
Общ азот // Total nitrogen	1,4 - 2,1%	- Устина // Ustina	
Катрани // Tar	35,40 - 39,60 mg/cig.	Година / Year Площ (хил ха) // Area Количество (тон) // Quantity	
Никотин(дим) // Nicotine (smoke)	0,77 - 1,08 mg/cig	▶ 1998 1.12 1415	
		1999 0.74 1215	
		2000 0.42 446	
		2001 0.65 820	
		2002 0.51 592	

a) Chemical composition of tobacco and tobacco smoke

b) Time series on areas and quantities

Figure 3. Tobacco_BG database: Varieties cube

Legislation cube of Tobacco_BG database integrates texts containing legal and normative information on tobacco production in Bulgaria. Figure 4 shows the opening of text document “Chapter 7. Agriculture” through database techniques and software system Adobe Acrobat.

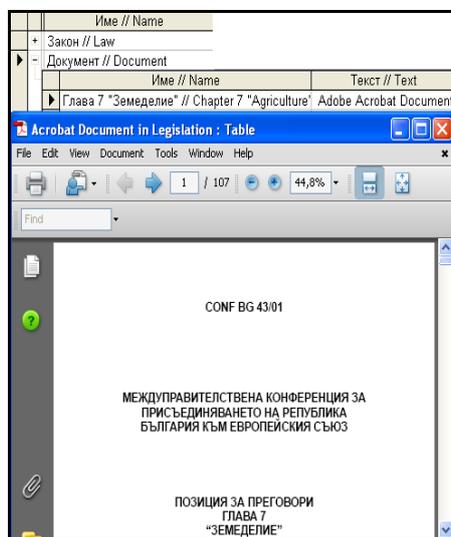


Figure 4. Tobacco_BG database: Legislation cube

Tobacco_BG database focuses on different user groups:

- Scientists and lecturers in the fields of biology, chemistry and phytopharmacy with interests in tobacco crop research and teaching activities
- Agronomists and managers responsible for cultivation, crop protection measures and quality of produced tobacco
- Experts and specialists from governmental institutions working with scientific and legal information on tobacco production.

Multidimensional database, based on open hierarchical structures provides developers with the opportunity to use it as a platform for building other databases. This can be done as follows:

- Building a database on tobacco production for other country will require only minimal changes in metadata of dimension tables and insertion of concrete data in fact tables
- Integrating expert information on several diverse crops can be realized by adding new dimensions and/or multiplication of hierarchies.

2. Conclusion

Hierarchically structured multidimensional database is systematic solution for integrating multi-aspect expert information on tobacco production. This is open solution regarding the number of dimensions and updates of stored information, thus giving opportunity to be used as a platform for building databases on diverse crops in other countries. The created database on tobacco production in Bulgaria enables users to work in friendly environment which helps them to avoid difficulties associated with finding information and operating with different files and formats. This solution increases expertise of scientists, teachers, experts and managers in their academic work and in the process of decision-making.

References

- [1] Lenzerini, M. (2002). Data integration: a theoretical perspective, Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 233-246.
- [2] Cali, A., Calvaneze, D., Giacomo, G., & Lenzerini, M. (2013). Data Integration under Integrity Constraints, Seminal Contributions to Information Systems Engineering, *Springer*, 335-352.
- [3] Hearst M. (1998). Trends & Controversies: Information integration. *IEEE Intelligent Systems* 13(5), 12-24.
- [4] Jensen, C., S., Pedersen, T., B., & Thomsen, Ch. (2010). Multidimensional Databases and Data Warehousing, *Synthesis Lectures on Data Management*, 2(1), 2010, 1-111.
- [5] Guo, X Ma, L. Su, K. & Sun, Y. (2014). Research and Design of Multi-Source Heterogeneous Information Integration Platform for Metadata Service, *Applied Mechanics and Materials*, 1485, 513-517.
- [6] Shubert, L., Jeffery, K., & Neidecker-Lutz, B. (2010). The Feature of Cloud Computing – Opportunities for European Cloud Computing Beyond 2010, *Expert Group Report – European Commission*, 1–66.
- [7] Wade, T. (2006). A to Z GIS: An Illustrated Dictionary of Geographic Information Systems, ESRI press.
- [8] Rana, Y., Lia, X. Lua, L., & Lib, Z. (2012). Large-scale land cover mapping with the integration of multi-source information based on the Dempster–Shafer theory pages, *Int. J. Geographical Information Science*, 26(1), 169-191.
- [9] Dimova, D. (2007). Onkov, K. An algorithm for automated creation of a PC database storing related text objects, *Journal of Information Technologies and Control*, 5(2), 48-52.

- [10] Dimova, D. (2010). An Algorithmic Solution for Management of Related Text Objects with Application in Phytopharmacy, *Serdica Journal of Computing*, 4, 487-504.
- [11] Ministry of Agriculture in Bulgaria. List of authorized plant protection products for placing on the market and use, registered fertilizers, soil improvers and nutrition substrates, 2013.
- [12] Dimitrov, A., Dimanov, D., & Kotchev, I. (2000). A new model for tobacco pest control in Bulgaria, Information bulletin "2000 CORESTA CONGRESS LISBON, PORTUGAL", section P9, p 129.