

ИНТЕГРИРАНИ СРЕДИ ЗА СЪБИРАНЕ, ОРГАНИЗАЦИЯ И ОБРАБОТКА НА ГОЛЕМИ ОБЕМИ ОТ ДАННИ

ПРОЕКТ 2019-ФЕЕА-01

Тема на проекта:
Интегрирани среди за събиране, организация и обработка на големи обеми от данни

Ръководител:
проф. д-р Цветозар Георгиев

Работен колектив:
доц. д-р Милко Маринов, доц. д-р Ирена Вълвова, доц. д-р Георги Георгиев, доц. д-р Светлана Стефанова, доц. д-р Силян Арсов, доц. д-р Анелия Иванова, доц. д-р Галина Иванова, гл. ас. д-р Орлин Томов, гл. ас. д-р Йордан Калмуков, гл. ас. д-р Елица Ибраимова, гл. ас. д-р Николай Костадинов, гл. ас. д-р Лъчезар Йорданов, гл. ас. д-р Цанко Големанов, гл. ас. д-р Емелия Големанова, гл. ас. д-р Хованес Аваки, 11 докторанти и постдокторанти, и 10 студенти

Адрес: 7017 Русе, ул. "Студентска" 8, Русенски университет "Ангел Кънчев"
Тел.: 062 - 888 711
E-mail: TGeorgiev@ecs.uni-ruse.bg

Целта на проекта е да се разработи инструментално средство, чрез което да се изследва производителността при запис и четене на данни в СУБД, поддържащи следните модели на данните: релационен, от типа ключ-стойност, колоноориентиран, документен.

Основни задачи:

- Извършване на сравнителен анализ на софтуерни архитектури за реализация на инструментално средство за оценяване производителността на СУБД.
- Създаване на архитектурни модели на системи за работа с големи обеми от данни.
- Разработване на генератор на експериментални данни за различни модели.
- Разработване на инструментално средство за оценяване производителността на СУБД.
- Тестване и оптимизиране на разработките.
- Популяризиране на разработката чрез участие с презентации, доклади, статии и експозити в научни семинари, конференции, изложби и др.

Основни резултати:

- Разработване на генератор на експериментални данни за различни модели.
- Разработване на инструментално средство за оценяване производителността на СУБД.

Публикации:

- Маринов, М. Модели на данни в нерелационните системи. Монография. Бряг Медиа Груп ЕООД, Русе, 2019, стр. 120, ISBN 978-619-90584-4-2.
- Вълвова, И. Изследване на релационни бази от данни. Монография. Бряг Медиа Груп ЕООД, Русе, 2019, стр. 148, ISBN 978-619-90584-3-5.
- Marinov, M., I. Valova. Component Interaction In Distributed Knowledge-Based Systems.// TEM Journal, 2019, No 8(3), pp. 721-727, ISSN 2217-8309. (SJIR rank: 0.148 /2018, Scopus)

Други:

- Още тридесет и три научни публикации, от които петнадесет индексирани от Scopus или Web of Science.

АНОТАЦИЯ

В резултат на човешката дейност в редица области от живота се генерират голямо количество несвързани документи, които обикновено се съхраняват само на хартия. Хартиеният носител не позволява да се извършва бързо и лесно търсене, а от там и статистически анализ на натрупаните през годините данни. Дори и в случай на електронни документи, писаната на взаиморазбиращи между тях силно ограничавана гледната точка на експертите и не позволява да се провеждат достатъчно задълбочени и широкообхватни анализи и прогнози. В резултат крайните изводи и заключения на анализаторите често са непълни или недостатъчно прецизни, което води след себе си и до неточни и/или неефективни стратегии за управление на процесите от съответните области.

Съвременните информационни и комуникационни технологии биха могли да имат съществен принос в решаването на този проблем. Преминването от хартиени към електронни документи ще позволи тяхното последващо автоматизирано обработване. Ако върху тях се приложат различни методи и алгоритми за семантичен анализ, то несвързаните на пръв поглед документи биха представили една много по-пълна и богата информация за процесите, за които се отнасят. Последващият семантичен и статистически анализ на тази информация ще позволи автоматизирано да се правят изводи и да се откриват тенденции, а от там и да се планират и оптимизират съответните процеси.

Данните в една система трябва да бъдат съхранявани и да бъдат достъпни във всеки един момент за тези, които се нуждаят от тях. За да могат данните да бъдат използвани, те трябва да бъдат обработени по подходящ начин с цел извлечане на необходимата информация, която ще се използва в процеса на вземане на решения. Данните се получават от т. нар. "оперативни системи – системи за вземане на оперативни решения" (On-Line Transaction processing (OLTP) systems – системи за обработка на трансакции в реално време), които извършват специфични задачи по отношение събиране и съхраняване на данните. При опит за анализ на такава информация, се открояват следните проблеми: едни и същи данни могат да бъдат съхранени в различните подсистеми; еднакви по своята природа същности са дефинирани по различен начин; данните са неподходящи; данните се изменят с времето.

Тези проблеми могат да се решат чрез интегриране на данните. За да се улесни процеса на вземане на решения, е разработана нова технология, по-сложна от СУБД, наречена хранилище (склад) за данни (Data Warehouse). Хранилището на данни може да бъде дефинирано като инструментално средство за подпомагане процеса за вземане на решения, което събира своите данни от оперативните бази от данни и различни външни източници, преобразува ги в информация и я прави достъпна за тези, които вземат решения, в консолидиран и логичен формат. Много модерно е в областта на големите обеми от данни да се използват нерелационни бази данни (NoSQL), които предоставят механизъм за съхранение и възстановяване на данни, като използват свободен съгласуван модел.

Днес все по-актуални стават различни варианти на интелигентни решения за търсене на изображения, за автоматизация на дома, офиса или другите сфери от ежедневието ни. При разработването и използването на такива решения се изисква събиране на огромно количество данни от различни сензори, съхраняването им, анализ и обработка, с цел прогнозиране и интелигентно управление.

Проектът разширява и задълбочава изследователската дейност в областта на интегрираните среди за събиране, организация и обработка на големи обеми от данни, извършвана в Русенския университет.

PROJECT 2019-FEEA-01

Project title:
Integrated environments for collecting, organizing, and processing big data

Project director:
Prof. Dr. Tsvetozar Georgiev

Project team:
Assoc. Prof. Dr. Milko Marinov, Assoc. Prof. Dr. Irena Valova, Assoc. Prof. Dr. Georgi Georgiev, Assoc. Prof. Dr. Svetlana Stefanova, Assoc. Prof. Dr. Silyan Arsov, Assoc. Prof. Dr. Anelia Ivanova, Assoc. Prof. Dr. Galina Ivanova, Ch. Assist. Prof. Dr. Orlin Tomov, Ch. Assist. Prof. Dr. Yordan Kalimukov, Ch. Assist. Prof. Dr. Elitsa Ibraymova, Ch. Assist. Prof. Dr. Nikolay Kostadinov, Ch. Assist. Prof. Dr. Lachezar Yordanov, Ch. Assist. Prof. Dr. Tsanko Golemanov, Ch. Assist. Prof. Dr. Emilia Golemanova, Ch. Assist. Prof. Dr. Hovanes Avakyan, 11 PhD students and post-doctoral students, and 10 students

Address: University of Ruse, 8 Studentska str., 7017 Ruse, Bulgaria
Phone: +359 82 - 888 711
E-mail: TGeorgiev@ecs.uni-ruse.bg

Project objective:
The aim of the project is to develop an instrumental tool to study the performance of recording and reading data in database management system supporting the following data models: relational, key-value, column-oriented, documentary.

Main activities:

- Performing a comparative analysis of software architectures for the realization of a tool for estimating the productivity of database management system.
- Creating architectural models of systems for handling big data.
- Development of an experimental data generator for different models.
- Development of a tool for estimating the productivity of the database management system.
- Testing and optimizing the developments.
- Promoting the project by participating with presentations, reports, articles and exhibits in scientific seminars, conferences, exhibitions, etc.

Main outcomes:

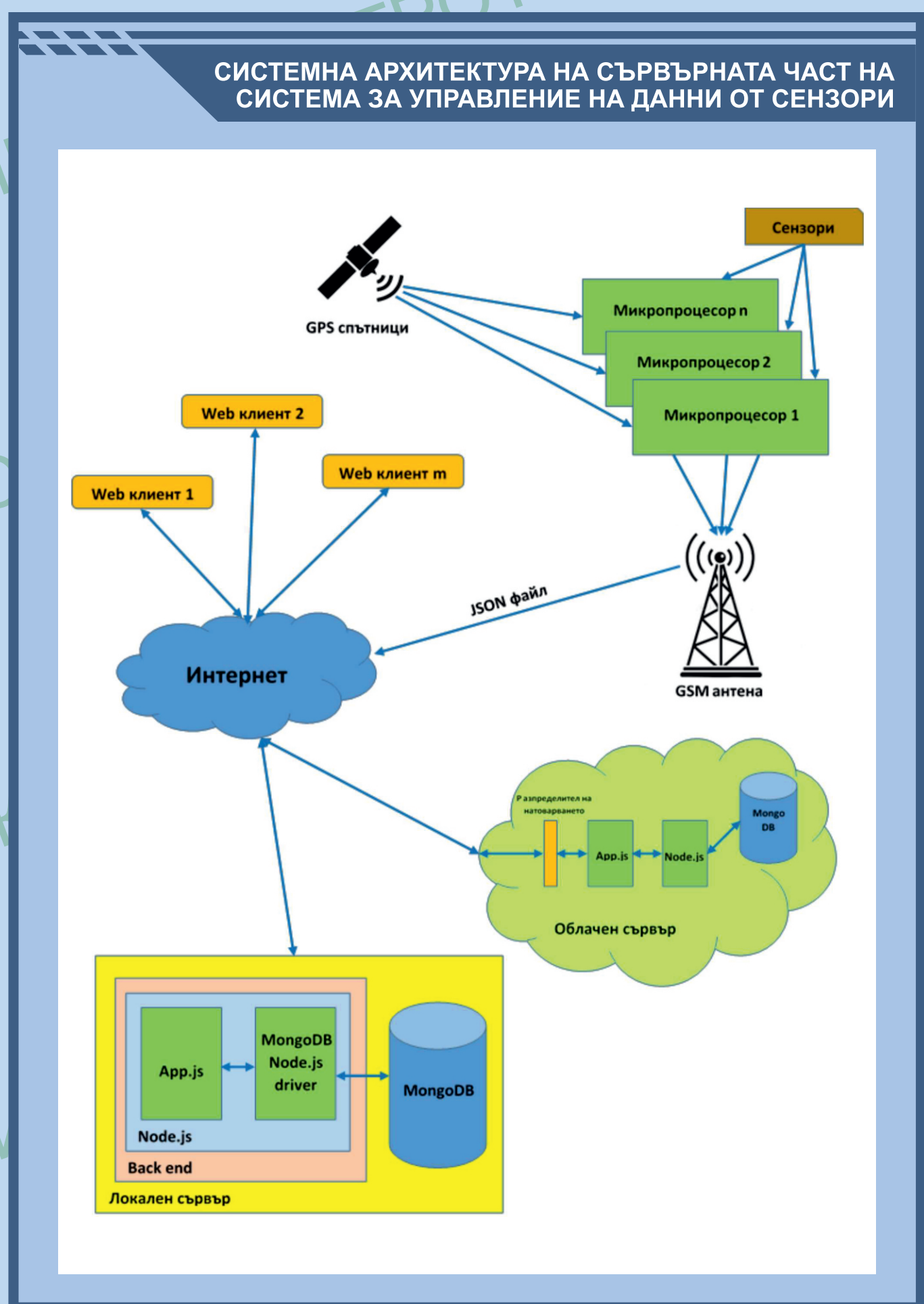
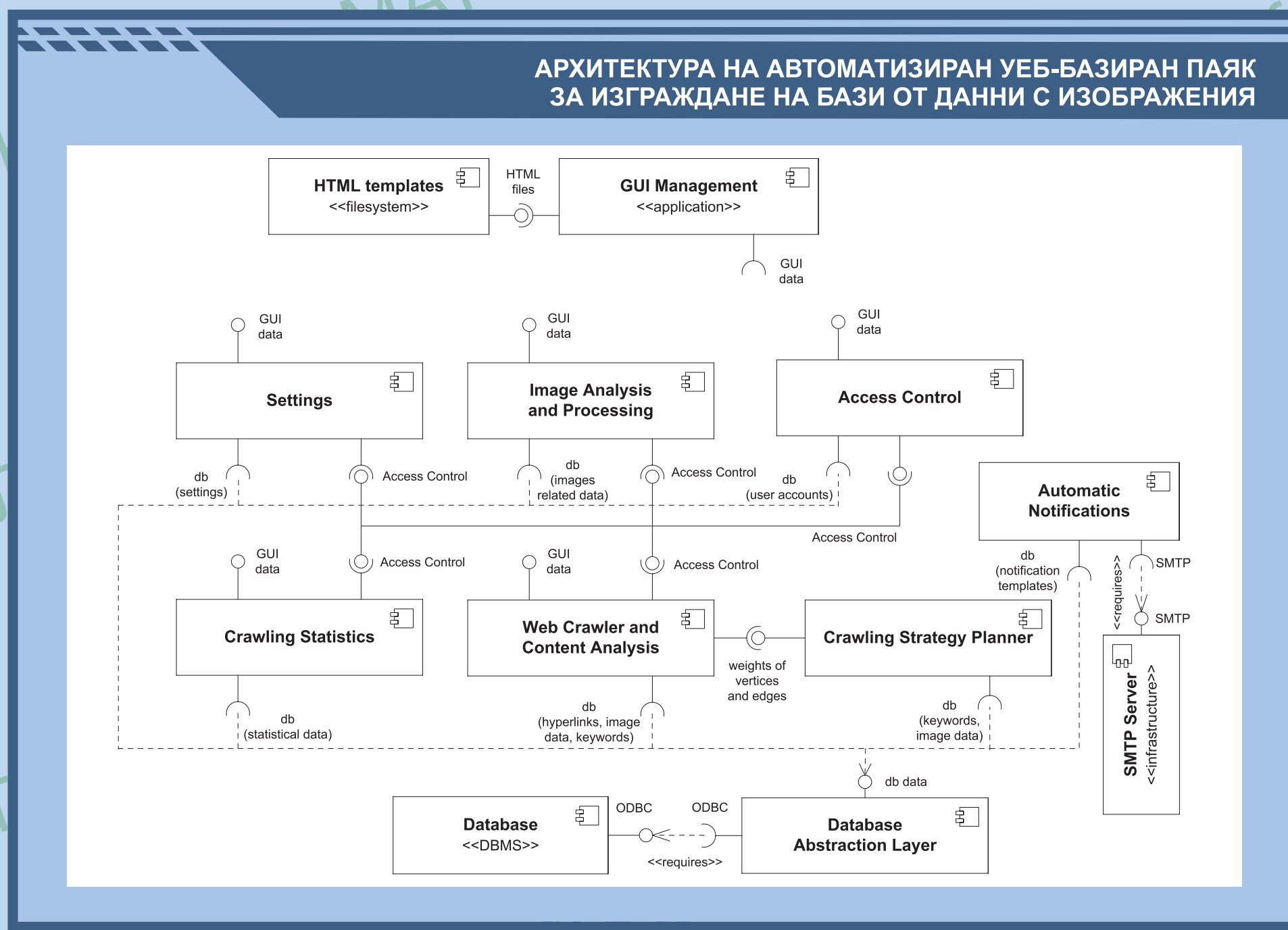
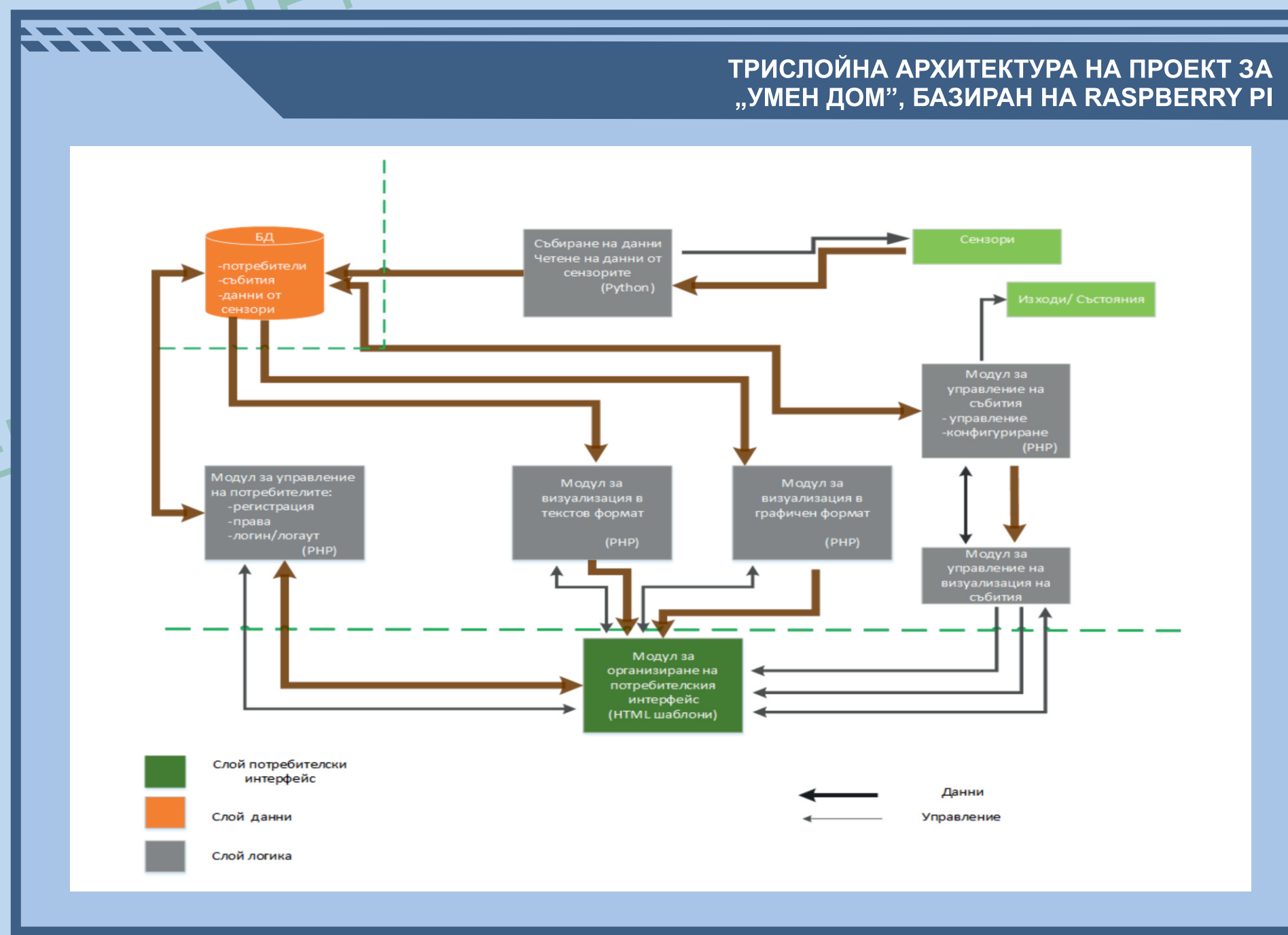
- Development of an experimental data generator for different models.
- Development of a tool for estimating the productivity of the database management system.

Publications:

- Marinov, M. Data Models in Non-relational Systems. Monograph. Bryag Media Group Ltd., Ruse, 2019, p. 120, ISBN 978-619-90584-4-2.
- Valova, I. Investigation of Relational Databases. Monograph. Bryag Media Group Ltd., Ruse, 2019, p. 148, ISBN 978-619-90584-3-5.
- Marinov, M., I. Valova. Component Interaction in Distributed Knowledge-Based Systems.// TEM Journal, 2019, No 8(3), pp. 721-727, ISSN 2217-8309. (SJIR rank: 0.148 /2018, Scopus)

Others:

- Thirty-three more scientific publications, fifteen of which are indexed by Scopus or Web of Science.



СРАВНИТЕЛЕН АНАЛИЗ НА СИСТЕМИ ЗА ИЗВЛИЧАНЕ НА ИЗОБРАЖЕНИЯ НА БАЗА НА СЪДЪРЖАНИЕТО

	ACORN	FREE	FINUTE	IOBB	VORAL	LINE	YANDEX	PASTEC	IMFILE	ПРЕДЛАГАНА СИСТЕМА
Вид на системата	WEB	WEB	WEB	DESK TOP	WEB	LIB	WEB	LIB	WEB	DESK TOP
Достъпност	FREE	FREE	PAID/FREE	FREE	FREE	FREE	FREE	FREE	FREE	FREE
Входни данни	IMG/TXT	IMG	IMG	IMG	IMG	IMG	IMG	IMG	IMG	IMG
За академични изследвания	Y	Y	N	Y	N	N	Y	N	N	Y
Доставчи	N	N	Y	N	N	N	Y	Y	N	N
Активна разработваща система	Y	Y	Y	Y	Y	N	Y	N	Y	Y
Уб. демо или за използване	Y	Y	Y	N	Y	N	Y	N	Y	N
Основни формати	JPEG, PNG	JPEG, PNG	JPEG, PNG, GIF, TIFF	JPEG	JPEG, BMP, PNG	N/A	JPEG, GIF, PNG	N/A	JPEG, BMP, PNG, GIF, TIFF	JPEG, BMP, PNG
Програмен език	N/A	C++, Python	N/A	N/A	N/A	JAVA	N/A	N/A	N/A	C#, VS
Цена (на месец)	N/A	FREE	FREE/around \$200	FREE	FREE	FREE	FREE	FREE	FREE	FREE
Използване на цветни функции	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Използване на текстови функции	N	Y	N	N	Y	Y	Y	Y	Y	N
Използване на формати функции	N	N	N	N	N	N	Y	N	Y	N
Използване на обратна връзка	Y	Y	Y	N	Y	N/A	Y	N/A	Y	N
Индексирани изображения (в милиони)	22	N/A	35 500	N/A	2 221	N/A	10 000	N/A	97	N/A
Лиценз	Closed	Open	Closed	Closed	Closed	GPL	Closed	LGPL	Open	Closed

