

СОФТУЕРНИ ИНСТРУМЕНТИ ЗА ВИЗУАЛИЗАЦИЯ НА ДИНАМИЧНИ ИНФОРМАЦИОННИ РЕСУРСИ

Милен Близнаков, Георги Пашев, Георги Тотков

Пловдивски университет „Паисий Хилендарски“
milen.bliznakov@gmail.com, georgepashev@gmail.com,
totkov@uni-plovdiv.bg

Резюме: Статията анализира и прави описание на системи за визуализация на информационни ресурси по подходящ начин. Представени са различни инструменти за моделиране, визуализация и анализ на данни. Създадена е таблица с характеристики на системите и е направено сравнение, базирано на възможностите на софтуерните системи за: визуален анализ, моделиране на диаграми и процеси, достъпност на потребители с различни технически умения, цели и начини на използване.

Ключови думи: софтуерни инструменти, динамична визуализация на данни, моделиране, визуален анализ

1. Въведение

За да може информацията да се разбира правилно е от значение да се представи в удобен вид. За тази цел е необходимо изобразената информация да бъде представена в подходяща логическа структура чрез предварително дефинирани, общовалидни конвенции. В повечето случаи, интерпретацията на информацията е свързана с културната и обществената среда, а също и с образователната и познавателна способност на участниците в процеса. Визуализацията на данни и последващата им обработка има нужда от софтуерни инструменти за тяхната интерпретация, което означава, че е необходимо специално да се изследва проблема за представянето на данните и ресурсите по подходящ начин.

2. Визуализация на динамични информационни ресурси

Визуализацията на информация [1] е набор от технологии, които използват визуални изчисления за разширяване на човешкото познание с абстрактна информация. Тя помага да се ускори разбирането на механизмите в един свят на нарастващ обем информация. Информационните визуализации [3] се опитват ефективно да картографират променливите данни върху визуалните измерения, за да създадат графични изображения. Изключително предизвикателство е, че визуализираните данни идват от много различни източници, всеки от които притежава различни характеристики. Също така данните могат да бъдат малък набор от данни само с няколко записа или голям

набор от данни с хиляди параметри, което също е част от проблема. Визуализацията на данни е термин [4], който описва усилията да се помогне на хората да разберат значението на данните, като ги поставят във визуален контекст. Моделите, тенденциите и корелациите, които могат да останат незабелязани в текстовите данни, могат да бъдат разкрити и разпознати полесно със софтуер за визуализация на данни. Днешните инструменти за визуализация на данни излизат извън рамките на стандартните диаграми и графики, използвани в електронните таблици, като показват данните по сложни начини чрез: инфографика; циферблати и измервателни уреди; географски карти; светещи линии; цветни карти (heatmap) и графики; стълбовидни, кръгли и линейни диаграми. Изображенията могат да включват интерактивни възможности, позволяващи на потребителите да манипулират с данните или да правят заявки и анализи. Могат да бъдат включени и индикатори, предназначени да предупреждават потребителите: при актуализиране на данните или при предварително зададени условия.

Динамични информационни ресурси могат да бъдат: релационни и NoSQL БД, електронни таблици, Excel, CSV файлове, JSON, GeoJSON, HTML и уеб-страници, XML, текстови файлове, уеб-услуги, виртуални дискове и хранилища, GitHub, социални мрежи, аналитични данни и източници, различни облачни структури и др.

Част от методите и начините за представяне на динамични информационни ресурси и данни са: хистограми, стълбовидни, линейни, кръгови, разпръснати, точки, конусовидни диаграми; цветни и мисловни карти; различни видове дървета и мрежи; обикновени и крос таблици; информационни платна и панели; матрици; графи и др. Има изследователи [2], които описват над 90 таблици, графики и диаграми, класифицирани в „Периодична таблица на визуализационните методи“, където са представени конкретни примери и характерни особености на различните прийоми за визуализация.

2.1. Системи за моделиране и визуализация на динамични информационни ресурси

VAX [5] е уеб-базиран визуален графичен редактор, който дава възможност за създаване на персонализиран визуален език за моделиране. За описанието на основните компоненти се използва скриптов език (Javascript). Изграждането на визуалния език става чрез предварително дефиниране на схема, типове, групи и блок-компоненти. Типовете са част от схемата и са подобни на статичните типове, които се дефинират в езиците за програмиране като Java, C++, Scala. Типовете могат да бъдат свързвани и групирани в граф-възли наречени блок-компоненти. Компонентът е шаблон за реален възел от основната схема и съдържа: дефиниции на входни и изходни гнезда, определения на атрибути, тип параметри с граници на типа, заглавие, цвят и

група. На практика, чрез предварително дефинираните блокове в схемата, визуалния редактор предоставя възможност за моделиране и изпълнение на процеси като граф-възли. Като резултатът от изпълнението е рекурсивна оценка на свързаните компоненти в модела.

Modelio [6] е среда за моделиране с отворен код. Предоставя широка гама от съществуващи модули, базирани на разпространени стандарти и функционалности за разработчици на софтуер, анализатори, дизайнери, бизнес архитекти и системни архитекти, като могат да бъдат адаптирани и разработвани персонализирани потребителски модули за конкретни нужди. Основните вградени модули са: Дизайнер на Java, който поддържа UML2 и Java, генериране на Java код и Javadoc; XSD Designer за графично моделиране на XML схеми (XSD модели), трансформиране на UML диаграми в XSD модели и генериране на XSD документи от XML модел; WSDL Designer за графично моделиране на уеб услуги (WSDL модели) и генериране на WSDL код от WSDL модел; TOGAF-архитект със специализиран редактор за използване на стратегии в корпоративното моделиране и за TOGAF-каталог; SysML Architect, който поддържа езика за моделиране на системи – SysML; UML Testing Profile, който поддържа стандартния език за тестване на OMG, предоставяйки специфични за теста концепции, специални артефакти и шаблони за генериране на последователни диаграми; SoaML Designer с редактор за моделирането на архитектурата на SOA и генериране на модели за внедряване. Стандартите, които поддържа са UML2, BPMN2, XMI, MDA, SysML, TOGAF, SoaML, UML Testing Profile. Има вградени функционалности като XMI-функция за импортиране и експортиране; публикуване на създадените модели в HTML формат; поддръжка на скриптовия език на Jython (интегриран Python с Java платформа), който се използва за онлайн заявки, малки скриптове и макро дефиниции.

Plotly [7] развиват и поддържат продукта Chart Studio, който дава възможност за създаване на графики, диаграми и различни таблични представяния. Анализът на данните и тяхното импортиране става чрез разпространени езици за програмиране и софтуерни продукти: JavaScript, Python, R, Matlab, Jupyter, Excel. Вградените библиотеки за визуализация и онлайн инструментите позволяват създаването на достъпни потребителски графики чрез уеб-базиран „drag-and-drop“ интерфейс. Предлага свързване с различни релационни БД (MySQL, PostgreSQL, RedShift, Oracle, Spark SQL/Hive, MsSQL, SQLite, Athena) и NoSQL хранилища (AWS S3, ElasticSearch, Apache Drill/Parquet, Google BigQuery, Apache Impala, data.world). Допълнително вече създадените графики могат да бъдат експортирани в различни формати: PDF, PNG, SVG, EPS, JPEG, HTML, JSON.

R е език [8] и среда за статистически изчисления и графики. Това е проект, разработен от Bell Laboratories от екип начело с Джон Чембърс. R може да се

разглежда като модификация на езика S [9]. R осигурява голямо разнообразие от статистически (линейно и нелинейно моделиране, класически статистически тестове, анализ на времеви серии, класификация, клъстеризация и др.) и графични техники. Основната разлика е: S е средство за избор за изследване в статистическата методология, а R осигурява софтуера с отворен код за участие в тази дейност. Една от силните страни на R е лекотата, с която могат да бъдат създадени добре проектирани графики за качествени публикации, включително математически символи и формули, където е необходимо. Специално внимание по подразбиране е обърнато на малките дизайнерски решения в графиката, като потребителят запазва пълен контрол върху дизайна. R е наличен като свободен софтуер и работи на голямо разнообразие от UNIX платформи (включително FreeBSD и Linux), Windows и MacOS. Включва: ефективни инструменти за обработка и съхранение на данни; пакет от оператори за изчисления на масиви, по-специално матрици; голямо, съгласувано, интегрирано събиране на междинни инструменти за анализ на данните; графични средства за анализ на данни и показване на екрана или на хартиен носител; добре развит, прост и ефективен език за програмиране, който включва условности, вериги, рекурсивни функции, дефинирани от потребителя, и възможности за вход и изход.

DataHero [10] е софтуерен инструмент за автоматизиране на бизнес информацията, за създаване на графики, отчети и информационни панели с вградена „drag-and-drop“ функционалност. Позволява събирането на данни от различни облачни услуги от външни източници и създаването на графики и таблични изгледи, без да са необходими технически умения. Възможните връзки към външни системи и източници на данни са: Google Drive, Google Sheets, Google AdWords, Google Analytics, Smartsheet, Dropbox, Microsoft OneDrive, Keen IO, Marketo, LinkedIn, Facebook, GitHub, Shopify, Salesforce, QuickBooks, Excel, CSV файлове и др. Предлага инструменти за филтриране и сегментиране с избор за различни типове графики и персонализирани цветови палитри, диаграми (линейни, кръгови, стълбови, блокови и др.), информационни и таблични, конфигурирани изгледи. Има възможност за експортиране на създадените диаграми като изображения, PDF файлове, или да бъдат вмъкнати като графики в уеб-сайтове и блогове на външни източници.

Tableau [11] включва софтуерни продукти за визуализация и анализ на многомерни масиви с данни. Предлага функционалност за чертаене на карти с географска дължина и ширина чрез извличане на данни от специализирани файлове (GeoJSON). Има възможности за осъществяване на връзки с релационни бази от данни, облачни бази от данни, електронни таблици и др., които след обработка могат да се визуализират под формата на: различни типове лентови, кръгови, стълбови диаграми и графики; крос таблици и матрици; картограми; хистограми; цветни и дървовидни карти; различни типове

дървета, графи и мрежи; маркиращи таблици; текстови таблици и др. Предлага възможности за инструменти и продукти с различно предназначение и цели (лично и професионално ползване, за организации, за бизнес процеси), между които Desktop, Server, Online и Data prep (платени версии) и Reader и Public (безплатни версии). Визията на Tableau Software е да представят „поетично“ данните, а основното мото на продукта е „да помага на хората да видят и разберат данните“. Комуникацията в софтуера е чрез иновация наречена VizQL [12], език за визуални заявки, преобразуващ „drag-and-drop“ действията в заявки за данни, което позволява на потребителите бързо да интерпретират данните.

Qlik [13] създава софтуерни инструменти за обработка на „сурови“ данни и трансформирането им в информация, готова за анализ, която всеки може да проучи. Мисията на Qlik е да създаде научно-ерудитан свят, в който всеки може да използва данните, за да разреши най-трудните си проблеми. Предлага: различни видове диаграми – линейни, комбинирани, кръгови, стълбови, грид, блокови, разпръснати (scatter), радиални (radar); обикновени, комбинирани и крос таблици; полета за настройки; различни агрегиращи функции върху данните; потребителски персонализирани обекти; импорт от релационни БД, текстови файлове и електронни таблици; манипулация чрез SQL-заявки; възможност за персонализирани скриптове и др. Графиките могат да бъдат експортирани в изображения, PDF и Excel. Предлага два инструмента за анализ на данни: QlikView и Qlik Sense. Разликата между двата продукта се състои основно в следното [14]: QlikView е за „структуриран, направляващ“ анализ, а Qlik Sense е за „самообслужване“ или собствени визуализации. QlikView е фокусиран върху създаването на инструментални табла за разработване, тяхното внедряване и предоставяне до крайния потребител по структуриран начин. Потребителят е в състояние да прегледа и анализира данните в съответствие с начина, по който е структурирано приложението, като се ръководи от процеса чрез използване на дизайна и оформлението. Qlik Sense е фокусиран върху крайния потребител, който управлява собствения си анализ по-скоро в „модел на самообслужване“. Qlik Sense дава възможност на крайния потребител да създава свои персонализирани визуализации, графики и обекти, което му позволява да управлява визуализацията на данните по собствен начин.

D3 е библиотека [15] с отворен код на JavaScript за създаване на динамични, интерактивни визуализации на данни в уеб браузъра. D3 има вградени функции за обработка на голям обем от данни и последващо представяне в подходящ формат. Използва широко прилаганите SVG, HTML5 и CSS стандарти. Предварително създадените функции дават възможност за създаване на обекти с: избор на подходящи елементи, добавяне на динамични ефекти или инструкции към тях и манипулиране на стила и визията им. Може да работи с големи масиви от данни, които могат да бъдат в различни формати JSON, CSV,

GeoJSON. Изходът за генериране на резултатите е в графични или текстови диаграми.

Платформата KNIME Analytics Platform [16] е софтуер с отворен код за създаване на информационни приложения за данни и услуги. Интуитивна и с допълнителни модули за интеграция на нови разширения, KNIME дава възможност за интерпретация на данни и проектиране на работни потоци и компоненти за многократна употреба. Има възможност за връзка с различни източници на данни: текстови формати (CSV, PDF, XLS, JSON, XML и др.); изображения; документи; мрежи (networks); бази данни; складове за данни; достъпване и извличане на данни от източници като Twitter, AWS S3, Google Sheets и Azure. Експортът на готовите графики и отчети може да бъде в PDF, Powerpoint, а съхранението на данните – в общи файлови формати или бази данни. Чрез вградените инструменти могат да се създават визуални работни потоци от различни възли, включвайки скриптове от Python, R, машинно обучение или конектори към Apache Spark. Има вградени допълнителни функции за агрегиране, сортиране, филтриране, извличане на статистики за средна стойност, прилагане на тестове за валидиране на хипотеза, анализ на корелация и стандартно отклонение и др. Предоставя възможности за изграждане на модели за машинно обучение, като използва различни алгоритми, включително дълбоки невронни мрежи (deep learning), методи базирани на дървета (tree-based methods) и логистична регресия (logistic regression). Визуализацията на данните може да бъде персонализирана според потребителските нужди като: бар-диаграма; диаграма на разпръскване; паралелни координати; мрежов граф; цветна карта (heatmap) и др.

Oracle Reports [17] е комплекс от програмни инструменти за генериране на справки и отчети на компания Oracle Corporation. Има възможност за достъп до различни информационни ресурси релационни БД, XML и текстови файлове. Справките могат да бъдат публикувани в различни формати (HTML, XML, PDF, Excel, RTF, PostScript и др.). Част от функционалностите и инструментите на Oracle Reports са: инструмент за валидиране на входните параметри, редактор на заявки с графично представяне на SQL-изразите; графичен редактор за оформление на отчетите; инструменти за генериране на динамични уеб-страници в зависимост от данните; създаване на отчети въз основа на възникване на определени събития в базата данни и др.

JasperSoft е група от софтуерни инструменти на TIBCO [18] за създаване и манипулация на справки, които осигуряват отчетност и анализи от различни източници на данни. Има възможност за връзки с голям набор от източници: релационни и NoSQL бази данни, CSV, Java Bean, Excel и собствени формати. JasperSoft поддържа комерсиална и свободна версия с ограничени функционалности. Предоставя REST и SOAP базирани уеб услуги и HTTP APIs за опростена интеграция с външни приложения. Част от функционалностите,

които предлага са: добавяне на изображения и създаване на под-справки; интерактивни функции, включващи „drop-and-down“ менюта, филтриране, форматиране, анимирани диаграми; крос-таблицы и аналитични изгледи; уеб-базиран дизайнер на доклади с възможности за „влачене и пускане“. Извежданите справки са във формати HTML, XLS, XLSX, PDF, CSV, DOCX, RTF.

Crystal Reports [19] е система на компанията SAP SE и осигурява интерактивна манипулация на справки и визуализация на различни информационни ресурси. Източниците на данни, с които работи, са релационни БД, електронни таблици, крос-таблицы и собствени хранилища. Функционалностите, които предлага продуктът са: добавяне на изображения и създаване на под-справки; създаване на сложни формули и персонализирани функции за обработка на данни; създаване и обработка на крос-таблицы, интерактивни панели за задаване на полета и входни параметри; инструменти за създаване на достъпно съдържание на отчети за хора с увреждания и др. Справките могат да се експортират в различни формати като: Word, Excel, PDF и актуализирани XML формати, RTF, TXT.

SQL Server Reporting Services (SSRS) [20] е софтуерна система на корпорация Microsoft за генериране на отчети на базата на услуги и вградени сървърни инструменти. Предоставя възможности за създаване, публикуване и управление на отчети, след което ги доставя на потребителите по различни начини, независимо дали ги преглеждат в уеб браузър, на тяхното мобилно устройство или като имейл. Може да се използва за създаване и обработка на различни интерактивни и печатни отчети, като осигурява интерфейс към среди за разработка (Visual Studio), SQL бази данни и възможност за интегрирани инструменти за форматиране на SQL отчети по сложни начини. „Report Builder“ е инструмент за потребители без технически и програмистки умения за форматиране на SQL отчети с по-малка сложност. По този начин, вместо да пише код, потребителят може да използва „drag and drop“ функционалност за повечето аспекти на SSRS доклада. Отчетите могат да се изтеглят в различни формати, включително DOC, Excel, PDF, CSV, XML, TIFF и други графични формати, HTML уеб архив и др. Предоставя възможност за мобилни отчети с адаптивно оформление, което се адаптира към различни устройства и ги визуализира по подходящ начин на мобилни устройства, с различна разделителна способност.

Таблица 1 представя описаните дотук софтуерни системи, според възможностите им за: моделиране; визуален анализ на данни; интерактивна визуализация; различни начини и методи за представяне на информацията; типове потребители, ползващи системата; експорт на данните в различни видове формати. Всеки от критериите за сравнение е оценен с точки от 0 до 3.

Таблица 1: Класификация на системи за моделиране и визуализация

Софт. система	ВАД	ВМ	ИВ	ВП	ДП	ВЕД
VAX	0	2	2	1	2	0
Modelio	0	3	2	2	3	1
KNIME	2	3	3	1	3	2
D3	2	0	2	0	3	1
Tableau	3	1	3	2	3	2
Qlik	3	1	3	1 – 3	3	2
ChartStudio	3	1	2	2	2	3
R	1	0	2	0	2	1
DataHero	3	1	2	2	3	2
Oracle Reports	2	2	3	2	2	3
JasperSoft	1	2	3	1	2	3
CrystalReports	1	2	3	2	3	3
SQL Server Reporting Services	1	2	3	2	3	3

Легенда:

0 – най-ниско ниво или не притежава функционалността; **1** – ниско ниво; **2** – средно ниво; **3** – най-високо ниво.

ВАД – възможност за анализ на данни (0 – 3);

ВМ – възможност за моделиране (с доп.познания – 0, без познания – 3);

ИВ – интерактивни визуализации: дали продукта има интегрирани функционалности за манипулация на информационните ресурси (0 – 3);

ВП – вид потребител (за разработчици – 0, без техн. умения – 3);

ДП – динамично представяне: различни начини и методи за представяне на информацията (0 – 3);

ВЕД – възможност за експорт на данни във формати (0 – 3).

2.2. Изводи

Представените програмни инструменти са сравнително популярни, но са една малка част от продуктите, които са налични в областта на софтуерните инструменти за моделиране и визуализация на информационни ресурси. Всички продукти са обединени от условието: да предлагат интерактивни визуализации и възможност да съдържат функционалности, които могат да описват процеси от различни предметни области. Част от програмните инструменти са предназначени по-скоро за обикновени потребители без познания, но има и такива, които са единствено и само за разработчици.

Заклучение

Статията описва софтуерни инструменти за моделиране, анализ на данни и динамично представяне на информацията от различни информационни източници. Представена е класификация на програмните продукти по различни характеристики. Подобни софтуерни инструменти успешно се употребяват за визуализация на данни в различни приложни области като: метеорология, молекулярна биология, химия, архитектура, геология, музика, география, за описание на различни видове модели и др. Обзорът на представените инструменти ще бъде от помощ за разработчиците, които тепърва навлизат в областта на визуализацията на информация като представя сравнителна характеристика на описаните програмни продукти и дава насоки за различните цели, за които могат да бъдат използвани.

Благодарности

Работата е подкрепена от проект МУ19-ФТФ-001 „Интелигентен анализ на данни за подобряване на резултатите от обучението“, към Фонд „Научни изследвания“ на ПУ „Паисий Хилендарски“.

Литература

1. Card, S., Information visualization, The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Lawrence Erlbaum Assoc Inc, p.539-542, ISBN: 978-0-8058-5870-9, 2008.
2. Lengler R., Eppler M. Towards A Periodic Table of Visualization Methods for Management. IASTED Proceedings of the Conference on Graphics and Visualization in Engineering, Clearwater, Florida, USA., 2007
3. Gee, A., M. Yu, H. Li, G. Grinstein, Dynamic and Interactive Dimensional Anchors for Spring-Based Visualizations. Technical Report, Computer Science, University of Massachusetts Lowell, 2005.
4. Rouse, M., Definition data visualization, <https://searchbusinessanalytics.techtarget.com/definition/data-visualization>, посл. посетен на 02.03.2019 год.
5. Ayupov VI., Visual Scripting is Easy Now, <https://aveic.github.io/vax/usage.html>, посл. посетен на 07.04.2019 год.
6. Modelio, The open source modeling environment, <https://www.modelio.org/about-modelio/features.html>, посл. посетен на 06.04.2019 год.
7. Plotly, Chart Studio - sophisticated editor for charts, <https://plot.ly/online-chart-maker/>, посл. посетен на 19.03.2019 год.
8. The R Project for Statistical Computing, <https://www.r-project.org/>, посл. посетен на 03.04.2019 год.
9. Chambers, J., Facets of R, R Journal, Volume 1(1) (May 2009), pp. 5-8., ISSN 2073-4859

10. DataHero, The fastest easiest way to automate business insights, <https://datahero.com/>, посл. посетен на 29.03.2019 год.
11. Tableau Software, <https://www.tableau.com/about/mission>, посл. посетен на 15.01.2019 год.
12. Jones, B., Communicating Data with Tableau, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2014, ISBN: 978-1-449-37202-6
13. Qlik, Lead with data -management and analytics platform, <https://www.qlik.com/us/>, посл. посетен на 28.03.2019 год.
14. Prima Consulting, Blog - QlikView vs. Qlik Sense, <https://primaconsulting.co.uk/news/2018/3/26/qlikview-vs-qliksense>, посл. посетен на 28.03.2019 год.
15. D3 – Data Driven Documents, <https://github.com/d3/d3/wiki>, посл. посетен на 15.11.2018 год.
16. KNIME Analytics Platform, <https://www.knime.com/knime-software/knime-analytics-platform>, посл. посетен на 27.03.2019 год.
17. Oracle Reports, https://docs.oracle.com/cd/E12839_01/bi.11111/b32121/pbr_intro001.htm#RSPUB23202, посл. посетен на 19.11.2018 год.
18. TIBCO Jaspersoft, <https://community.jaspersoft.com/documentation?Version=49176>, посл. посетен на 20.11.2018 год.
19. SAP Crystal Reports – create powerful, richly formatted and dynamic reports, <https://www.crystalreports.com/reports/>, посл. посетен на 15.11.2018 год.
20. SQL Server Reporting Services, Microsoft, What is SQL Server Reporting Services, <https://docs.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-2017>, посл. посетен на 04.04.2019 год.

SOFTWARE TOOLS FOR VISUALIZING DYNAMIC INFORMATION RESOURCES

Milen Bliznakov, George Pashev, George Totkov

Abstract: *The work presents systems for the visualization of information in an appropriate way. Describes templates for modeling templates and visual analysis. A classification of the systems has been developed and the comparison is based on the capabilities of software systems for data analysis, modeling systems, user level, goals and ways of use.*

Keywords: *software tools, dynamic visualization, modeling, visual analysis*