

ГРАФИЧНО ПРЕДСТАВЯНЕ НА ДАННИ ЧРЕЗ ИНСТРУМЕНТИ С ОТВОРЕН КОД

Иван Казанджиев¹, Николай Янев², Кънчо Иванов³

¹Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail

²Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail niki@mgu.bg

³Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail

РЕЗЮМЕ. В статията се разглежда теорията за качество на данните и в частност тяхното графично представяне. Направен е обзор на няколко инструмента използващи различни съвременни програмни технологии за презентиране на данни. За повишаване на качеството на данните бяха избрани инструменти с отворен код, поради тяхната добра функционалност, достъпност и възможност за лесно надграждане.

GRAPHICAL REPRESENTATION OF DATA WITH OPEN-SOURCE SOFTWARE

Иван Казанджиев¹, Николай Янев², Кънчо Иванов³

¹Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail

²Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail niki@mgu.bg

³Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail

ABSTRACT In this paper we delve into the theory of Data Quality and in particular – graphical presentation of data. We have reviewed a few programming instruments that are using different modern programming technologies capable of visually presenting data. For the improvement of Data Quality were chosen programming products that are open-source because of their good functionality, accessibility, and easy modification.

Качество на данните

Напоследък интензивно се развива теорията Data Quality (фиг. 1). Основна задача на тази теория е разработката на методи и средства за анализ и за подобряването на качеството на данните.



Фиг. 1. Data Quality - интеграционна платформа.

Осигуряване на Data Quality (DQ) е процес на проверка на надеждността и ефективността на данни. Качеството на

данните е и степента, до която информацията изпълнява определени изисквания. Поддържането на качеството на данните изисква обработването им и периодичното им пречистване. Обикновено това включва анализиране, актуализиране, стандартизиране и избягване на дублиране на данни, за да създаде единен изглед на данните, включително и от разнородни системи.

Степента на детайлизацията или общото представяне на информация е избор от страна на потребителите по време на диалога и създава благоприятна работна среда с оглед разнообразните условия за взимане на решение. От друга страна подробната информация, която е представена чрез разнообразно оцветяване и специализирано сигнализиране създава удобство при работата на потребителите и им спестява времето за допълнителен анализ.

Подходи за разработване на приложение, удовлетворяващо Data Quality

Има три основни подхода, за да се направи оценка дали дадени данни са подходящи да се използват в определен контекст:

- Теоретичен - този подход има само четири свойства за постигане на качеството на данните (пълнота, еднозначност, значимост, вярност). Тези свойства не осигурява надежден модел за всички сфери на приложимост на DQ.

- Интуитивен — подход, който се базира на интуицията на признат експерт в сферата на DQ, който да определи необходимите свойства, които да притежават данните, за да са полезни за даденото им приложение.

- Емпиричен – при този подход техническите изисквания на клиентите и тяхната преценка служат за определяне на дадени данни като качествени.

основа на своите категоризации за качество на данните, добра комбинация за определянето на качеството е съчетаването на интуитивния и емпиричен подход.

Томас С. Редман – участник в развитието на теорията за Data Quality и експерт в сферата, предлага следните 4 категории с общо петнадесет свойства за определяне качеството на данните :

Като се има в предвид, че уважаваните и признати експерти в тази сфера използват теоретичния подход като Таблица 1.

Категория	Свойство	Определение
Същественост	Правдоподобност	“Степента, до която данните се разглежда като истина, са реални и надеждни.”
	Точност	“Степента, до която данните се разглежда като верни, надеждни и несъдържащи грешки.”
	Обективност	“Степента, до която данните са безпристрастни.”
	Репутация	“Степента, до която данните са достоверни и високо ценени заради източника си.”
Контекст	Добавена стойност	“Степента, до която данните са полезни и носят предимства при тяхното използване.”
	Уместност	“Степента, до която данните са приложими за текущата задача.”
	Своевременност	“Степента, до която данните са актуални за текущата задача.”
	Цялост	“Степента, до която данните са достатъчно задълбочени в сферата на текущата задача.”
	Подходящо количество данни	“Степента, до която обемът на данните е достатъчен.”
Презентация	Интерпретируемост	“Степента, до която данните са в подходящият език и единиците и дефинициите са ясно представени.”
	Леснота на възприемане	“Степента, до която данните са очистени от двусмислие и лесно се възприемат.”
	Презантационна последователност	“Степента, до която данните са винаги представяни в един и същи формат и са съвместими с предишни данни.”
	Кратко представяне	“Степента, до която данните са компактно представени и все пак точни и достатъчни.”
Достъпност	Достъпност	“Степента, до която данните са достъпни.”
	Сигурност	“Степента, до която достъпът до данните може да бъде ограничен.”

Както е видно от таблица 1 презентацията на данните е една от основните четири категории, определящи качеството на дадени данни. Структурирайки множеството числови стойности и представянето им с помощта на визуални методи (графики) спомага за по-бързо и лесно възприемане на представената ни информация.

Разработването на програмни продукти следвайки добрите практики препоръчвани по отношение на графичната визуализация на данните в Data Quality (таблица 1) води до лекота при възприемане на информацията от потребителите и съкращаване на техническото време за вземане на решение, което по същество създава условия за по-ефективна работа.

Инструменти за графично представяне на данни

За избор на най-подходящ набор от технологии и инструменти за създаване на графични визуализации, бяха проучени и анализирани няколко такива:

1. **Highcharts JS** – е JavaScript библиотека, която предоставя богато разнообразие от графични модели за представяне на данни. Информацията за изчертаването на графиките се подава в един от два формата – XML или JSON (текстови формати).

1.1. Част от възможностите на продукта Highcharts JS са:

- 1.1.1. Възможност за създаване на динамично изчертавани и интерактивни графики.

- 1.1.2. Множество настройки по визуалната част на графиките (анимиране на елементи от графиката, тип на графиката, цветовете, етикети, промяна на посоката на изписване на текст и други).
- 1.1.3. Възможност за принтиране и запазване на графиките като изображения.
- 1.2. **Предимства на Highcharts JS.**
Използва JavaScript – език за програмиране, който се изпълнява директно от брауъра и има отлични перспективи за развитие, HTML за структуриране на информацията, като за изчертаване на самите изображения се използва технологията SVG при повечето брауъри (тези които не поддържат SVG, използват Flash като алтернативен вариант).
- 1.3. **Минуси на Highcharts JS.**
Highcharts JS е платен продукт, когато се използва с комерсиална цел и лиценз за един програмист струва 360\$.
2. **Flot** – е JavaScript библиотека, която предоставя малко разнообразие от графични модели за представяне на данни. Информацията за изчертаването на графиките се подава JSON формата.
 - 2.1. **Част от възможностите на продукта Flot са:**
 - 2.1.1. Възможност за създаване на динамично изчертавани графики.
 - 2.1.2. Настройки по визуалната част на графиките (тип на графиката, цветовете, етикети и други).
 - 2.2. **Предимства на Flot.**
Използва както и Highcharts JS JavaScript, HTML и SVG. Безплатен продукт и за работата си изисква jQuery – друга JavaScript библиотека, също безплатна.
 - 2.3. **Минуси на Flot.**
Малък набор от настройки и липсва интерактивност на графиките.
3. **amCharts** – предлага също доста голяма гама от възможности за създаване на графики. Продукта използва JavaScript и XML или CSV файл за приемане на данните за представяне.
 - 3.1. **Част от възможностите на продукта amCharts са:**
 - 3.1.1. Възможност за създаване на динамични и интерактивни графики.
 - 3.1.2. Множество настройки по визуалната част на графиките (тип на графиката, цветовете и други). Като цяло възможностите за персонализация са сходни, но по-малко от тези на **Highcharts JS**.
 - 3.2. **Предимства на amCharts.**
Продуктът също използва JavaScript и HTML технологиите, но има и алтернативен вариант за работа с Flash технологията вместо JavaScript.
 - 3.3. **Минуси на amCharts.**
amCharts е платен продукт и лиценз за един програмист струва 99€. Има опция за безплатно използване с комерсиална цел, но изискването е да се включи хипервръзка към сайта на производителя в крайния продукт. Цената е по-ниска от тази на **Highcharts JS**, но и функционалността е по-ограничена.
4. **ChartDirector** – е PHP базирана система за изчертаване на графики. Графиките са статични картинки създадени чрез **gd** библиотеката на PHP. Имплементации на **ChartDirector** има в няколко други езика: Java, .NET, ASP, COM, VB, Perl, Python, Ruby, ColdFusion, C++.
 - 4.1. **Част от възможностите на продукта ChartDirector са:**
 - 4.1.1. Възможност за създаване на статични графики.
 - 4.1.2. Множество настройки по визуалната част на графиките (тип на графиката, цветовете, етикети и други).
 - 4.2. **Предимства на ChartDirector.**
Няма такива
 - 4.3. **Минуси на ChartDirector.**
ChartDirector е платен продукт и лиценз за един програмист струва 99\$.
5. **XML/SWF** – е продукт, който се възползва от Flash и XML технологиите, за да създаде интерактивни и динамични графики.
 - 5.1. **Част от възможностите на продукта XML/SWF са:**
 - 5.1.1. Възможност за създаване на динамични и интерактивни графики.
 - 5.1.2. Множество настройки по визуалната част на графиките (тип на графиката, цветовете, етикети, анимиране на елементи от графиката и други).
 - 5.2. **Предимства на XML/SWF.**
Що се отнася до нашият проект **XML/SWF** няма предимства пред **Open Flash Chart**.
 - 5.3. **Минуси на XML/SWF.**
XML/SWF е платен продукт и лиценз за един програмист струва 39\$.
(Не толкова минус колкото изискване)
XML/SWF се нуждае от добавката за който и да е от популярните интернет брауъри (Firefox, Internet Explorer, Chrome, Opera, Safari) - **Flash Player** - за да може да работи. При това Google Chrome включва автоматично тази добавка при инсталацията.
6. **Open Flash Chart** – е продукт, който използва Flash технологията за да трансформира скучните текстови числови данни във една динамична и интерактивна

графика. Информацията за изчертаването на графиките се подава в JSON формат.

6.1. Част от възможностите на продукта **Open Flash Chart** са:

- 6.1.1. Възможност за създаване на динамични и интерактивни графики.
- 6.1.2. Множество настройки по визуалната част на графиките (анимиране на елементи от графиката, тип на графиката, цветовете, етикети, промяна на посоката на изписване на текст и други).

6.2. Предимства на **Open Flash Chart**.

Продуктът предлага почти същата функционалност, каквато предлагат платените му конкуренти, но безплатно. Тъй като е лицензиран под GNU софтуерен лиценз - имаме право да модифицираме продукта и да го продадем т.е. използване с комерсиална цел.

6.3. Минуси на **Open Flash Chart**.

(Не толкова минус колкото изискване) **Open Flash Chart** се нуждае от добавката за който и да е от популярните интернет браузъри (Firefox, Internet Explorer, Chrome, Opera, Safari) - **Flash Player** - за да може да работи. При това Google Chrome включва автоматично тази добавка при инсталацията.

7. **Fusion Charts** – също използва Flash технологията за да анимира и представи по по-интересен начин информацията. Необходимите данни се подават в JSON формат.

7.1. Част от възможностите на продукта **Fusion Charts** са:

- 7.1.1. Възможност за създаване на динамични и интерактивни графики.
- 7.1.2. Много настройки по визуалната част на графиките (анимиране на елементи от графиката, тип на графиката, цветовете, етикети, промяна на посоката на изписване на текст и други).

7.2. Предимства на **Fusion Charts**.

Продуктът предлага почти същата функционалност, каквато предлагат **Open Flash Chart**. Също е лицензиран под GNU софтуерен лиценз - имаме право да модифицираме продукта и да го продадем т.е. използване с комерсиална цел.

7.3. Минуси на **Fusion Charts**.

(Не толкова минус колкото изискване) **Open Flash Chart** се нуждае от добавката за който и да е от популярните интернет браузъри (Firefox, Internet Explorer, Chrome, Opera, Safari) - **Flash Player** - за да може да работи. При това Google Chrome включва автоматично тази добавка при инсталацията. Спрямо **Open Flash Chart** минус, е че се използва по-старта версия на езика за програмиране във Flash (ActionScript). За в бъдеще от **Flash Player**-а може да отпадне възможността за изпълнение на код написан на старта версия.

Таблица 2.

Продукт	Промяна на графични елементи	Интерактивност	Обща Функционалност	Формат на данните	Цена за лиценз
Highcharts JS	Най-много на брой	Много добре реализирана	Отлична	JSON,XML	360\$
Flot	Малко на брой	Никаква	Слаба	JSON	Безплатен
amCharts	Много на брой	Добре реализирана	Много добра	XML,CSV	99€ (с опция за безплатен)
ChartDirector	Много на брой	Слабо реализирана	Добра	Директно от базата данни	99\$
XML/SWF	Много на брой	Добре реализирана	Много Добра	XML	39\$
Open Flash Chart	Много на брой	Много добре реализирана	Отлична	JSON	Безплатен
Fusion Charts	Много на брой	Добре реализирана	Много добра	XML	Безплатно

Спряхме се на продукта **Open Flash Chart** – тъй като напълно осигурява функционалността необходима за графичната визуализация на данните от справките в системата “VEZNA”, също така лицензът му е безплатен.

Заклучение

Продуктът **Open Flash Chart (OFC)** бе апробиран в най-голямото въгледобивно предприятие в Република България - “Мини Марица-изток” ЕАД. OFC беше използван за графично представяне на данни, което представлява част от уеб-базирана система, разработена за нуждите на “Мини Марица-изток” ЕАД. Потребителската

оценка от използването тази система, и в частност на графичната визуализация, е че качеството на данните - в смисъла на информативност, скорост на възприемане и удобство при работа са се повишили. Тази положителна потребителска оценка ни дава увереност да твърдим, че избраните подходи за постигане на Data Quality в комбинация с продукти с отворен код са дали много добър резултат.

Литература

Янев Н., К. Иванов, И. Казанджиев, Усъвършенстване на ИС с уеб базиран интерфейс на Програмния продукт за пренос и обработка на информация от ж.п. везна в

разтоварището на ТЕЦ Марица Изток 2, договор № МТ - 509/2009

Янев Н., К. Иванов, И. Казанджиев, Н. Халачев, Г. Георгиев, П. Костова, ПОВИШАВАНЕ КАЧЕСТВОТО НА ДАННИТЕ ЧРЕЗ ПРИЛАГАНЕ НА СИСТЕМАТА VEZNA, Годишник на МГУ „Св. Иван Рилски“, Том 53, Свितък IV, 2010

Redman T., “Data Quality: The Field Guide”, Digital Press 2001

Garvin D., “Competing on the Eight Dimensions of Quality”, Harvard Business Review, (November-December 1987), pp. 108-109.

Garvin D., “Managing Quality”; (New York: Free Press, 1987).

Batini C. M. Scannapieca – “Data Quality Concepts, Methodologies and Techniques”, Springer;