MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY

Portlet pro interaktivní vizualizaci dynamických dat

DIPLOMOVÁ PRÁCE

Tomáš Kratěna

Brno, podzim 2012

Prohlášení

Prohlašuji, že tato diplomová práce je mým původním autorským dílem, které jsem vypracoval samostatně. Všechny zdroje, prameny a literaturu, které jsem při vypracování používal nebo z nich čerpal, v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Vedoucí práce: RNDr. Radek Ošlejšek, Ph.D.

Shrnutí

Cílem práce je vytvořit portlet do systému Liferay, který umožní zobrazovat data z webových diskuzí pomocí různých typů grafů. Vycházíme z dat poskytnutých průmyslovým partnerem IBA CZ, s.r.o.

Klíčová slova

vizualizace dat, graf, chart, diskuze, časová řada, portál, portlet

Obsah

1 **Úvod** 1
  1.1 *Požadavky* 1
2 **Grafy** 3
  2.1 *Vývojový graf* 3
  2.2 *Srovnávací graf* 3
3 **Návrh** 4
  3.1 *Data* 4
  3.2 *Graf* 5
Literatura 7

Kapitola 1

Úvod

V rámci průzkumu a zpracování dat uživatelem je klíčová jejich vizualizace. Čím větší množství údajů je zpracováváno a čím složitější je jejich struktura, tím méně je člověk schopen bez technické pomoci mezi nimi najít nebo z nich odvodit zajímavé informace. Internet má potenciál obsahovat takové informace, jejich získání však vzhledem k objemu dat obvykle není jednoduché.

Pro širší možnosti interpretace je potřebná škála více jejich zobrazení. Pro různá zpracování mohou být vhodné různé vizualizace a pro některá může být užitečná možnost data reprezentovat více způsoby současně. Pro usnadnění práce by měly být vizualizace uživatelsky interaktivní a při současném použití by měly interagovat i mezi sebou.

Průmyslový partner má k dispozici strukturovaná data z webových diskuzí, která jsou představována hierarchií příspěvků. K příspěvkům jsou kromě základních údajů jako je čas vložení a autor přiřazeny také další údaje jako naměřená hodnota citového zabarvení diskuze, tzv. *sentiment* — nálada příspěvku vyjádřená číslem z intervalu od 1 (pozitivní) do 5 (negativní). [5]

V rámci diplomových prací, na něž tato navazuje, byly vyvinuty dvě vizualizace: *spectrogram* [3] a *diagram* [1], které mají podobu Java portletů v rámci portálové aplikace [6]. Obě slouží k zobrazení hierarchie diskuzních příspěvků, přičemž zohledňují hodnotu jejich sentimentu. Cílem této práce je najít vhodnou knihovnu pro vykreslování různých grafů a s její pomocí vytvořit další portlet, který umožní zobrazovat různé typy grafů nad daty získanými z diskuze, a integrovat jej do existující aplikace. Portlet bude sloužit k zobrazení časových řad pro identifikaci klíčových hráčů [5].

1.1 Požadavky

Cílem je uživateli umožnit do portálu umístit více instancí vytvořeného portletu a v každé dovolit výběr zobrazovaného grafu. Zobrazení má být interaktivní a dynamicky reagovat na změnu dat či parametrů provedenou v jiném portletu s použitím meziportletové komunikace.

Požadavkem je, aby byl návrh aplikace co nejobecnější. Má být navrženo obecné rozhraní představující graf včetně zobrazených dat a vizuálního stylu. Implementace pak bude na tomto rozhraní založena.

V návaznosti na předchozí práce má být vizualizace realizována kombinací technologií SVG a JavaScript [3]. Portlet bude integrován do existující portálové aplikace, která poběží

na portálovém serveru Liferay [4].

Je nutné, aby byl portlet použitelný v komerční sféře, tedy i použitá knihovna musí z licenčního pohledu splňovat tento požadavek.

Kapitola 2

Grafy

V této kapitole si představíme nejpoužívanější typy grafů podle jejich vizuální reprezentace. Dále se budeme snažit abstrahovat základní typy podle toho, jakou informaci předávají.

2.1 Vývojový graf

Základním typem je graf, který slouží k zobrazení vývoje hodnot jedné nebo více proměnných. Pro potřeby této práce mu říkejme *vývojový graf*. Typickou ukázkou je graf matematické funkce. Jde o vynesení hodnot proměnných v bodech, kde jsou definovány.

Primárním smyslem grafu je zobrazit vývoj a změny vývoje každé proměnné, tedy rozdíl hodnoty v bodě proti hodnotám v okolních bodech. Dalším informací, kterou graf předává v případě, že zobrazuje více proměnných, je jejich porovnání – jak srovnání hodnot v jednotlivých bodech, tak průběhů v celém rozsahu bodů.

Často se hodnoty jednotlivých proměnných zobrazují s jinými měřítky tak, aby se maxima a minima proměnných promítala na stejnou úroveň. To je užitečné, například když má některá proměnná nesrovnatelně menší hodnoty než jiná a zajímají nás změny v jejím vývoji – hodnoty proměnné jsou při společném měřítku stlačeny tak, že se informace o jejím vývoji vytrácí. Nezávislými měřítky se maximalizuje rozdíl hodnot ve vedlejších bodech stejné proměnné, čímž se informace o vývoji zpřesňuje. Nepříjemným vedlejším efektem je, že se znemožní srovnání proměnných vůči sobě. Proto nemusejí být měřítka nezávislá u všech proměnných.

2.2 Srovnávací graf

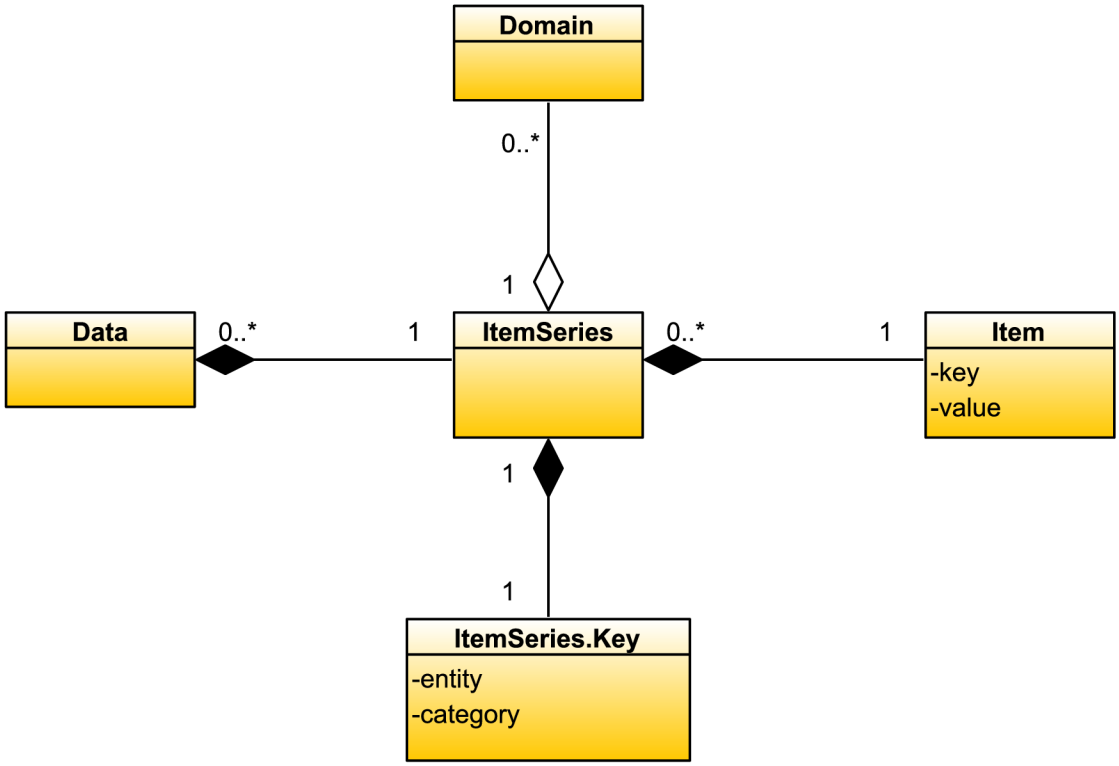
Dalším typem je graf, který porovnává několik hodnot. Pojmenujme jej *srovnávací graf*.

Kapitola 3

Návrh

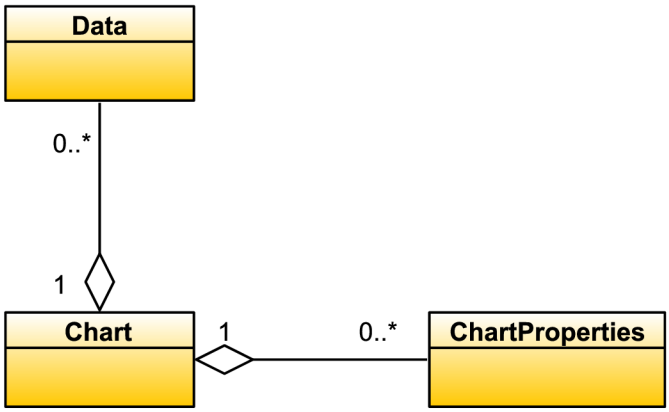
3.1 Data

Klíčovým požadavkem pro návrh modelu pro data vizualizovaná grafy byla obecnost. Ačkoliv různé typy grafů zobrazují jiné množství informací, které jsou jinak strukturované, naším cílem je mít jeden univerzální model. Z pohledu dat to znamená, že model musí být schopen poskytnout informace všem typům grafů.

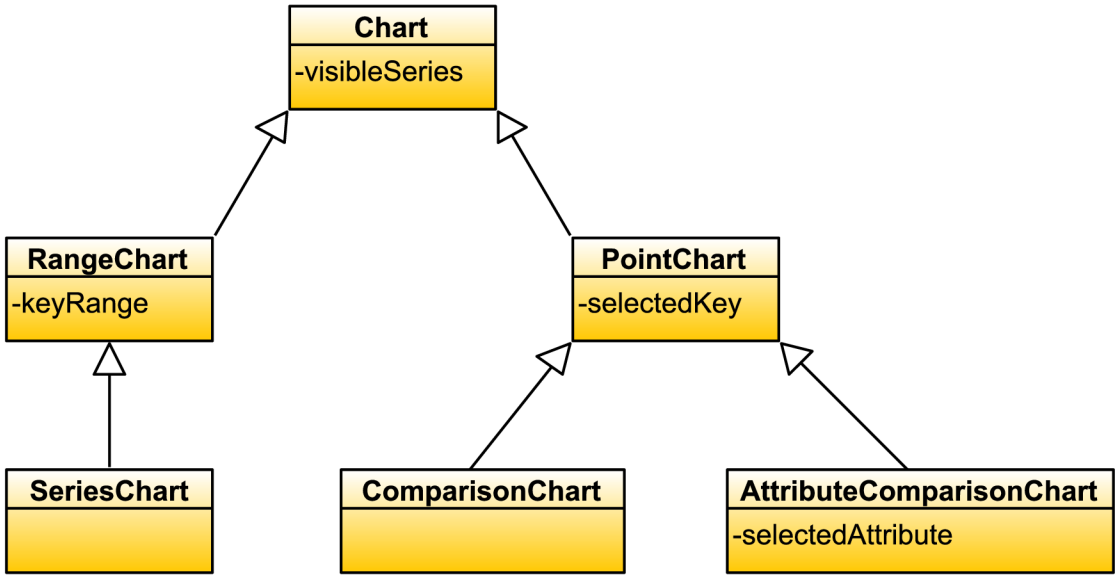


Obrázek 3.1: Diagram znázorňující strukturu dat.

3.2 Graf



Obrázek 3.2: Diagram znázorňující strukturu grafu.



Obrázek 3.3: Diagram znázorňující hierarchii grafů.

Literatura

[1] Dunčák, P.: *Vizualizace sentimentu diskusních vláken: diplomová práce*, Brno: Masarykova univerzita, Fakulta informatiky, jaro 2012. 1

[2] Gilbert, D.: *The JFreeChart Developer Guide*, Object Refinery Limited, February 9, 2007.

[3] Kedaj, Z.: *Spektogramy webových diskusí: diplomová práce*, Brno: Masarykova univerzita, Fakulta informatiky, podzim 2011. 1, 1.1

[4] Sezov, R.: *Liferay in action*, Manning Publications Co., 2011, ISBN 9781935182825. 1.1

[5] Pařilová, T.: *Vizualizace obsahu internetových diskuzí: diplomová práce*, Brno: Masarykova univerzita, Fakulta informatiky, 2012. 1

[6] Sarin, A.: *Portlets in Action*, Manning Publications Co., 2011, ISBN 9781935182542. 1