

Dynamisk Scrapeapplikation

– En paraplyapplikation för hästmarknaden



**LUNDS
UNIVERSITET**

Lunds Tekniska Högskola

LTH Ingenjörshögskolan vid Campus Helsingborg
MultimediaTeknik

Examensarbete:
Andreas Månsson
Gustav Ekberg

© Copyright Andreas Månsson, Gustav Ekberg

LTH Ingenjörshögskolan vid Campus Helsingborg
Lunds universitet
Box 882
251 08 Helsingborg

LTH School of Engineering
Lund University
Box 882
SE-251 08 Helsingborg
Sweden

Tryckt i Sverige
Media-Tryck
Biblioteksdirektionen
Lunds universitet
Lund 2008

Sammanfattning

Denna rapport beskriver examensarbetet vi gjort för LBS Science Park AB. Anledningen till att vi gjort just denna applikation var att LBS Science Park såg ett problem i den Internetbaserade hästannonsmarknaden. Det fanns ingen ledande annonssida, utan användare som skulle sälja eller köpa hästar fick använda sig av flera mindre sidor. Vår uppgift blev att undersöka marknaden och att skapa en applikation som samlar annonser från flera hemsidor på vår applikation.

Rapporten är indelad i tre huvuddelar, som belyser de juridiska, de tekniska och de innehållsmässiga frågorna. Svaren på dessa tre huvudkategorier har sedan legat som grund för vårt genomförande.

Detta projekt resulterade i en applikation vars huvudfunktion är fullt implementerad. Den presenterar annonser som är hämtade från andra applikationer och dessa går att söka igenom. Mängden annonser applikationen får in är ca 450 per dag.

Mervärdessfunktionerna, fråga experten och köparskolan är inte fullt implementerade än och kommer vidareutvecklas innan lanseringen.

Abstract

This report is the bachelor thesis we did for LBS Science Park AB. The reason we did this application was because LBS Science Park had identified a problem in the Internet based horseadmarket. There were no leading website, instead users who wanted to buy or sell a horse had to search through many smaller sites. We got the assignment to investigate the market and create an application that collects advertises from many sites and presents them in one place.

This report is divided into three main parts, in which we elucidate the legal, the technical and the content associated questions. The answers to these questions have later been the foundation of our implementation.

This project resulted in an application which main function is fully implemented. It presents advertises that's collected from other applications and it's possible to search through these. The amount of advertises the application collects per day is about 450. Other functions, ask the expert and school for buyers isn't fully implemented yet and will be implemented before the launch of the application.

Innehållsförteckning

1 Inledning	8
1.1 Bakgrund	8
1.2 Målsättning.....	8
1.3 Problemformulering.....	9
1.3.1 Huvudproblem.....	9
1.3.2 Delproblem.....	9
1.3.2.1 Juridiska frågor:.....	9
1.3.2.2 Tekniska frågor:.....	9
1.3.2.3 Innehållsrelaterade frågor:.....	9
2 Metodik.....	10
2.1 Inledning.....	10
2.2 Orddefinitioner.....	10
2.3 Förberedande undersökning	10
2.4 Analys	10
2.5 Kravspecifikation.....	11
2.6 Scraping	11
2.7 Utförande.....	11
3 Avgränsningar	12
4 Kravspecifikation	13
4.1 Systemkrav	13
4.1.1 Funktionella krav	13
4.1.2 Datakrav.....	13
4.1.3 Säkerhet.....	15
4.1.4 Innehåll	15
4.2 Gränssnitt mot hårdvara	16
4.2.1 Server	16
4.2.2 Klient	16
4.3 Projektkrav	16
4.3.1 Utvecklingsmiljö	16
4.3.2 Testkrav	17
4.3.3 Dokumentation.....	18
4.3.4 Leveranskrav	18
5 Juridiska frågor	19
5.1 Finns det någon information vi inte får ta?	19
5.2 Hur måste vi presentera källorna?	20
5.3 Vilken text och vilka bilder får vi ta från andra sidor?	20
6 Tekniska frågor	21
6.1 Hur fungerar scrapingtekniken?	21
6.1.1 Regular expressions (regex).....	21

6.2 Hur implementerar vi scrapingtekniken?	21
6.3 Hur löser vi att GlobalHorseShop ska finnas i olika språk?	22
6.4 Skall vi ha två separata tabeller, en för scrapingresultat och en för annonser som läggs upp direkt på GlobalHorseShop?	23
6.5 Hur ser vi till så att vi inte får två av samma annons?	23
6.6 Hur länge skall annonser få vara kvar på GlobalHorseShop?	23
6.7 Hur ofta uppdaterar vi informationen vi tar från andra databaser?	23
6.8 På vilket sätt ska annonsören kunna ladda upp bilder?	23
6.9 Hur löser vi bildscrapingen?	24
6.10 Vilka programmeringsspråk skall vi använda oss av?	24
6.10.1 Val av språk	24
7 Innehållsfrågor	25
7.1 Vilket material ska vi utgå ifrån när vi formar gränssnittsapplikationen?	25
7.1.1 Målgrupp	25
7.1.2 Undersökning	25
7.1.3 Frågorna	25
7.1.4 Resultat	25
7.2 Hur bestämmer vi gränssnittsapplikationen identitet?	27
7.2.1 Brus	27
7.2.2 Färgschema	29
7.2.3 Visuellt design	29
7.2.3.1 Design	29
Hur skall strukturen i gränssnittsapplikationen vara uppbyggd?	31
7.2.4 Navigation	31
7.2.5 Funktionalitet	31
7.2.6 Grundfunktionalitet	32
7.2.6.1 Annonssök	32
7.2.6.2 Användarsök	32
7.2.7 Mervärdes funktioner	32
7.2.7.1 Medlemskap	32
7.2.7.2 Bli medlem	32
7.2.7.3 Nyhetsbevakaren	33
7.2.8 Begränsningar	33
7.3 Hur presenterar vi bäst informationen för användaren?	33
8 Utförande/process	34
8.1 Metodik	34
8.2 Introduktion till Globalhorseshop	34
8.3 Framtagning av material	34

8.3.1 Vår första anblick på GlobalHorseShop	34
8.4 Utveckling av applikationen	35
8.4.1 Förstasidan	35
8.4.2 Annonsuppbyggnad	36
8.4.3 Scraping	36
9 Testning	37
9.1 Användartest 1	37
9.1.1 Bakgrund	37
9.1.2 Slutsatser	37
9.1.3 Åtgärder	37
9.2 Användartest 2	37
9.2.1 Bakgrund	37
9.2.2 Slutsatser	37
9.2.3 Åtgärder	37
9.3 Kontinuerlig testning	38
10 Resultat och diskussion	39
10.1 Utvärdering av förundersökning	39
10.2 Juridik	39
10.3 Utvärdering av applikationen	39
10.4 Problem	40
10.5 Olösta problem	40
10.6 Diskussion	41
10.7 Avslutning	42
11 Referenser	43
12 Bilagor	44
12.1 Enkät	44
12.2 Användartest 1	47
12.2.1 Scenario	47
12.2.2 Genomförande	47
12.3 Användartest 2	49
12.3.1 Scenario	49
12.3.2 Genomförande	49

1 Inledning

1.1 Bakgrund

Problemet med Internet idag är inte mängden information utan spridningen av den. Många marknader består av ett flertal mindre applikationer som har en liten del, det vill säga att det inte finns någon dominerande applikation för marknaden.

En av dessa marknader är hästannonsering. Här finns många mindre hemsidor där användare kan lägga upp annonser om hästförsäljning och relaterade affärer. De flesta existerande hemsidor har annonseringen som en biprodukt och koncentrerar sig mer på exempelvis nyheter.

En ide om hur man ska nå ut till hela marknaden är att skapa en såkallad ”paraplyapplikation”. En paraplyapplikation tar information från mindre applikationer som existerar på marknaden och presenterar den informationen på en samlad plats.

LBS Science Park AB är ett Varbergbaserat utvecklingsbolag. Bolaget ägs av Varbergs Sparbank AB, LBS Intressenter AB, Rickard Eriksson via sitt bolag Bokaberg AB och Robert Bengtsson via VästMedia Robert Bengtsson AB. Rickard Eriksson och Robert Bengtsson skapade Lunarstorm, som var världens första community.

Scraping är en typ av teknik som gör det möjligt för användare att spara ner information från en annan applikation till sin egen databas. Generellt sätt innebär det att en applikation skapas som går igenom koden på en sida, utefter det angivna mönster tar sedan applikationen ut den information som står i dessa mönster.

1.2 Målsättning

Att skapa en ekonomisk lönsam webbapplikation som samlar den utspridda hästmarknaden i nordn på ett och samma ställe. Där tanken är att alla användare tillslut ska skapa annonser direkt på GlobalHorseShop

1.3 Problemformulering

1.3.1 Huvudproblem

Att undersöka möjligheterna för att skapa en webbapplikation som samlar hästannonser på en och samma plats. Sedan skall vi med hjälp av resultaten av undersökningarna skapa applikationen.

1.3.2 Delproblem

Våra delproblem kan delas in i tre kategorier: de juridiska frågorna, de tekniska frågorna och innehållsrelaterade frågor. Problemen är rangordnade efter prioritering för varje kategori.

1.3.2.1 Juridiska frågor:

- Finns det någon information som vi inte får ta/använda?
- Hur måste vi presentera källorna?
- Vilken text får vi ta från andra sidor?
- Får vi ta andra sidors bilder?
- Hur skiljer sig lagarna i de olika länderna?

1.3.2.2 Tekniska frågor:

- Hur fungerar scrapingtekniken?
- Hur implementerar vi scrapingtekniken?
- Hur löser vi att GlobalHorseShop ska finnas i olika språk?
- Skall vi ha två separata tabeller, en för scrapingresultat och en för annonser som läggs upp direkt på GlobalHorseShop?
- Hur ser vi till så att vi inte får två av samma annons?
- Hur länge skall annonser få vara kvar på GlobalHorseShop?
- Hur ofta uppdaterar vi informationen vi tar från andra databaser?
- På vilket sätt ska annonsören kunna ladda upp bilder?
- Hur löser vi bildscrapingen?
- Vilka programmeringsspråk skall vi använda oss av?

1.3.2.3 Innehållsrelaterade frågor:

- Vilket material ska vi utgå ifrån när vi formar gränssnittsapplikationen?
- Hur bestämmer vi gränssnittsapplikationen identitet?
- Hur skall strukturen i gränssnittsapplikationen vara uppbyggd?
- Hur presenterar vi bäst informationen för användaren?

2 Metodik

2.1 Inledning

Vi har valt att delvis arbeta i en iterativ projektform inspirerad av SCRUM, och delvis arbeta sekventiellt.

I början av projektets gång, under undersökningarna kommer vi dock inte att arbeta efter denna modell eftersom där är etapper som behöver klaras av i en viss ordning. Det är inte förrän vi börjar utveckla applikationen som vi kommer att börja arbeta efter SCRUM.

SCRUM metodiken innebär att vi kommer att sätta nya mål under kortare perioder som vi kommer att arbeta med full kapacitet under just den tiden. Efter att varje nytt delmål är avklarat kommer vi att bestämma ett nytt inför nästa sprint.

2.2 Orddefinitioner

Definitioner på ord som används ofta i rapporten:

Användare: Person som har för avsikt att besöka sidan för olika ändamål.

Annonsör: En användare som har av avsikt att lägga upp en annons.

Scrapa: När vi scrapar en sida menar vi att vi med hjälp av Umbrellamotorn tar information från sidan och sedan lägger in i vår databas.

Insticksprogram: är ett program som körs utan att installeras, som en del av ett annat program. Ett annat ord för insticksprogram är plug-in.

2.3 Förberedande undersökning

Vi skall undersöka de juridiska möjligheterna/begränsningarna genom att ta kontakt med LBS Science Parks juridiska ombud.

När vi har fått svar på de juridiska frågorna måste vi undersöka de tekniska förutsättningarna, detta kommer vi att göra genom en dialog med LBS Science Park där de definierar vad vi skall använda för verktyg för att skapa applikationen.

Vi undersöker målgruppen genom att ta kontakt med personer vi fått rekommenderade av LBS Science Park och med hjälp av kontakter vi skaffat oss.

2.4 Analys

Baserat på informationen från undersökningen av målgruppen kan vi nu analysera resultat och skriva målgruppsanalysen.

I detta skede kommer vi att börja arbeta iterativt.

2.5 Kravspecifikation

Med all denna information och slutsatser som vi kommit fram till kommer vi nu att forma kravspecifikationen, detta kommer framöver att definiera vårt projekt och fungera som målsättning.

Nu kommer vi att skapa en visuell design som kommer att utforma sidan.

2.6 Scraping

Nästa steg är att vi måste lära oss att använda scrapingtekniken. Detta kommer att ske genom en kurs hos LBS SciencePark.

2.7 Utförande

Med denna information som vi har samlat ihop, kan vi börja skapa själva applikationen.

Vi börjar med att skapa ett skal baserat på den visuella designen.

Därefter kommer vi att skapa scrapingfunktionerna som vi sedan kommer att kunna applicera på det skalet.

Vi kommer sedan att skapa kontosystemet och annonsering och applicera det på skalet.

Sökfunktionen kommer nu att skapas, sökformuläret och

resultatpresentationen kommer vara det huvudsakliga i detta skede.

Säkerheten kommer att testas och vi skall undersöka de buggar som eventuellt finns och korrigera dem.

3 Avgränsningar

LBS ScienceParks mål med denna applikation är att den skall bli global, det vill säga att sidor skall scrapas från hela världen och den skall finnas i de flesta språk. Detta är dock något vi har valt att inte implementera. Vi har istället valt att skapa en dynamisk sida som enkelt kan anpassas för flera språk och olika typer av annonser.

Vad det gäller det juridiska frågorna har vi valt att överlåta det juridiska ansvaret på Science Park, då de har dragit de slutsatser som GlobalHorseShop grundas på.

Att användaren skall kunna ladda upp filmer är något vi valt att inte implementera, men tanken var att vi skulle skapa en guide om hur man laddar upp filmer på youtube och sedan länkar till dem från annonsen.

LBS SciencePark hade även som önskemål att vi skulle implementera ett släktträd där användare skulle kunna lägga upp sina hästar. Detta valde vi att inte implementera utan kommer att ligga som en del av vidareutvecklingen. Eftersom detta inte blir aktuellt förrän applikationen har ett stort antal användare.

4 Kravspecifikation

4.1 Systemkrav

4.1.1 Funktionella krav

4.1.1.1 Det ska finnas en registreringssida

4.1.1.2 Det ska finnas konto

4.1.1.2.1 Det ska gå att logga in

4.1.1.2.2 Det ska gå att logga ut

4.1.1.2.3 Det ska gå att ändra lösenord

4.1.1.3 Konto skall krävas för

4.1.1.3.1 Lägg upp en annons

4.1.1.3.2 Ändra i en annons

4.1.1.3.3 Ta bort en annons

4.1.1.4 Konto skall ej krävas för att titta på annonser

4.1.2 Datakrav

4.1.2.1 När en användare registrerar sig ska följande data registreras

4.1.2.1.1 Namn (efternamn, förnamn)

4.1.2.1.2 Adress

4.1.2.1.3 Telefonnummer

4.1.2.1.3 E-post

4.1.2.2 Om en användare begär nytt lösenord skickas det till given e-post

4.1.2.3 När en annons läggs upp på vår sida ska följande registreras

4.1.2.3.1 E-post

4.1.2.3.2 Mobilnummer

4.1.2.3.3 Hemsida

4.1.2.3.4 Publiceringsdag

4.1.2.3.5 Län/provins

4.1.2.3.6 Postnummer

4.1.2.3.7 Ort

4.1.2.3.8 Land

4.1.2.3.9 Häst

4.1.2.3.9.1 Young Rider

4.1.2.3.9.2 Senior

4.1.2.3.10 Ponny

4.1.2.3.10.1 A

4.1.2.3.10.2 B

4.1.2.3.10.3 C

4.1.2.3.10.4 D

4.1.2.3.11 Ras

4.1.2.3.12 Kön

4.1.2.3.13 Fader

- 4.1.2.3.14 Moder
- 4.1.2.3.15 Höjd
- 4.1.2.3.16 Färg
- 4.1.2.3.17 Inriktning
 - 4.1.2.3.17.1 Dressyr
 - 4.1.2.3.17.2 Fälttävlan
 - 4.1.2.3.17.3 Western
 - 4.1.2.3.17.4 Övrigt
- 4.1.2.3.18 SVRF nummer
- 4.1.2.3.19 FEI Nummer
- 4.1.2.3.20 Utbildning
- 4.1.2.3.21 Värt att veta
- 4.1.2.3.22 Tävling
 - 4.1.2.3.22.1 Lokal
 - 4.1.2.3.22.2 Regional
 - 4.1.2.3.22.3 Nationell
 - 4.1.2.3.22.4 Internationell
- 4.1.2.3.23 Pris
- 4.1.2.3.24 Fritext
- 4.1.2.3.25 Säljare
 - 4.1.2.3.25.1 Kontaktinfo
- 4.1.2.3.26 Datum/tid
- 4.1.2.4 När en annons läggs upp kan användaren välja att registrera
 - 4.1.2.4.1 Film
 - 4.1.2.4.2 Bild
 - 4.1.2.4.3 Pdf
- 4.1.2.5 Annonser ska ha en tidsbegränsning
 - 4.1.2.5.1 Tidsbegränsningen ska vara en månad.
 - 4.1.2.5.2 Efter tidsbegränsningen flyttas annonsen till ett arkiv
- 4.1.2.6 När en användare anpassar sitt sökformulär till defaultsök
 - 4.1.2.6.1 Om användaren är medlem kommer formuläret sparas mot användaren
 - 4.1.2.6.2 Om användaren ej är medlem kommer formuläret sparas till cookies
- 4.1.2.7 Scraping
 - 4.1.2.7.1 Scrapingen skall schemalagt utföras en gång var 5 minut
 - 4.1.2.7.2 Annonser ska hämtas från andra databaser
 - 4.1.2.7.3 Annonser ska inte hämtas mer än en gång
 - 4.1.2.7.4 Annonser ska sparas i en databas
 - 4.1.2.7.5 Följande information ska hämtas från en annons om informationen är tillgänglig för annonsen
 - 4.1.2.7.5.1 Namn
 - 4.1.2.7.5.2 Datum/tid

- 4.1.2.7.5.3 Kön
- 4.1.2.7.5.4 Mankhöjd
- 4.1.2.7.5.5 Ras
- 4.1.2.7.5.6 Fritext
- 4.1.2.7.5.7 Säljare
- 4.1.2.7.5.8 Kontaktinfo
- 4.1.2.7.5.9 Ort/land
- 4.1.2.7.5.10 Pris
- 4.1.2.7.5.11 Bild
- 4.1.2.8 Det ska finnas en databas för sidans annonser
- 4.1.2.9 Det ska finnas en databas för scrapade annonser
- 4.1.2.10 Det skall finnas en databas där information om alla besökare lagras
- 4.1.2.11 Varje besökares land skall registreras
- 4.1.2.12 Varje besökares webbläsare skall registreras
- 4.1.3 Säkerhet**
 - 4.1.3.1 Lösenord ska krypteras
 - 4.1.3.2 GHS har rättigheter att ta bort annonser
 - 4.1.3.4 GHS har rättigheter att ta bort konto
- 4.1.4 Innehåll**
 - 4.1.4.1 Det ska finnas en meny
 - 4.1.4.2 Det ska finnas reklam på sidan
 - 4.1.4.3 Det ska finnas kändisprofiler
 - 4.1.4.4 Det ska finnas möjlighet att ladda upp tävlingsresultat genom pdf.
 - 4.1.4.5 Det ska finnas en startsida
 - 4.1.4.5.1 Det ska finnas en nyhetsfeed
 - 4.1.4.5.2 Det ska finnas ett sökfält på annonser
 - 4.1.4.5.3 Det ska finnas ”thumbnails” för annonser
 - 4.1.4.5.4 Det ska finnas en ”senaste annonser” funktion
 - 4.1.4.6 Det ska finnas en söksida
 - 4.1.4.6.1 Det skall finnas ett dynamiskt sökformulär
 - 4.1.4.6.2 Formuläret ska ha följande kategorier:
 - 4.1.4.6.2.1 Fritext
 - 4.1.4.6.2.2 Ras
 - 4.1.4.6.2.3 Pris
 - 4.1.4.6.2.4 Mankhöjd
 - 4.1.4.6.2.5 Hästinriktning/Tävlingstyp
 - 4.1.4.6.2.6 Säljare
 - 4.1.4.6.2.7 Färg
 - 4.1.4.6.2.8 Ort/land
 - 4.1.4.6.2.9 Hästtyp

- 4.1.4.6.2.10 Namn
- 4.1.4.6.2.11 Kön
- 4.1.4.6.3 All sökinformation skall gå att lägga till och ta bort för användaren
- 4.1.4.6.4 Söksidan ska ha en starta sökning knapp
- 4.1.4.6.5 Det ska finnas en reset knapp
- 4.1.4.6.6 Det ska finnas en tipsa en vän knapp
- 4.1.4.7 Det ska finnas en annonshanterings sida
 - 4.1.4.7.1 Det ska finnas en profil för användaren
 - 4.1.4.7.2 Det ska gå att lägga till annons
 - 4.1.4.7.3 Det ska gå att ta bort annons
 - 4.1.4.7.4 Det ska gå att ändra annons
- 4.1.4.8 Det ska finnas en sida för säljare för inloggade användare
 - 4.1.4.8.1 Det ska finnas en översikt över användarens annonser
 - 4.1.4.8.1 Det ska gå att gå till skapa annons
 - 4.1.4.8.1 Det ska gå att ta bort annonser
 - 4.1.4.8.1 Det ska gå att ändra annonser
- 4.1.4.9 En support sida
 - 4.1.4.9.1 Det ska finnas kontaktinformation
 - 4.1.4.9.2 Det ska finnas FAQ
 - 4.1.4.9.3 Det ska finnas information om företaget

4.2 Gränssnitt mot hårdvara

4.2.1 Server

- 4.2.1.1 Linux Ubuntu 7.10
- 4.2.1.2 Apache
- 4.2.1.3 MySQL
- 4.2.1.4 PHP 5.0
- 4.2.1.5 Applikationen skall vara XHTML 1.0 certifierad
- 4.2.1.6 Applikationen skall vara CSS 2.1 certifierad

4.2.2 Klient

- 4.2.2.1 Skärmupplösning bör vara minst 1024x768 pixlar
- 4.2.2.2 Klienten måste ha java-script support
- 4.2.2.3 Klienten bör ha cookies support
- 4.2.2.4 Browsern ska klara av XHTML 1.0
- 4.2.2.5 Browsern ska klara av CSS 2.1

4.3 Projektkrav

4.3.1 Utvecklingsmiljö

- 4.3.1.1 Applikations utvecklas med:
 - 4.3.1.1.1 Text-editor
 - 4.3.1.1.2 Microsoft SQL Server management

4.3.2 Testkrav

4.3.2.1 Applikation ska användartestas av två personer från målgrupp.

4.3.2.2 Applikationen ska testas under utvecklingen

4.3.2.2.1 Efter varje större funktion i applikationen.

4.3.2.2.2 Varje vecka.

4.3.3 Dokumentation

4.3.3.1 Utgått

4.3.3.2 Det skall finnas versionshistorik av applikationen

4.3.3.3 Det ska finnas övergripande hjälptexter i applikationen vid:

4.3.3.3.1 Vid registrering av annonsör

4.3.3.3.2 Vid registrering av annons

4.3.3.3.2.1 Extra hjälptext för att lägga upp en film

4.3.3.3.2.2 Extra hjälptext för att lägga upp en pdf

4.3.3.3.3 Vid inloggning

4.3.4 Leveranskrav

4.3.4.1 Applikation anses färdig vid leverans till Science Park

5 Juridiska frågor

Att scrapa information från andra applikationer medför vissa juridiska frågor. Här följer de diskussioner vi hade med LBS SciencePark angående dessa juridiska frågor.

LBS Science Park ansvarar för alla juridiska punkter angående vårt projekt och det är upp till dem att undersöka lagarna gentemot en release av applikationen till allmänheten.

5.1 Finns det någon information vi inte får ta?

Att scrapa text är något som LBS SciencePark har gjort tidigare i andra applikationer. De har då undersökt detta djupgående och det visade sig att det inte fanns några lagar som konkret täcker detta område. Den lag som nämner något om detta verkar dock tyda på att det vi gör är lagligt:

URLr1 § 4: Den som översatt eller bearbetat ett verk eller överfört det till annan litteratur- eller konststart har upphovsrätt till verket i denna gestalt, men han äger icke förfoga däröver i strid mot upphovsrätten till originalverket. ’

Har någon i fri anslutning till ett verk åstadkommit ett nytt och självständigt verk, är hans upphovsrätt ej beroende av rätten till originalverket.[1]

Något som dock alltid måste lämnas är en källa, exempelvis en länk till originalartikeln.

Att scrapa bilder var något nytt. En tanke var att skicka ut e-post till alla annonsörer och få tillstånd att använda bilder i vår applikation. Detta förslag hade dock tagit ofantligt mycket tid och förmodligen väldigt dålig svarsfrekvens. Det hade inte fungerat i vår applikation för i genomsnitt får vi in ungefär 450 annonser per dag och alla annonser i andras applikationer har inte med e-post för annonsören. En annan lösning var att bara ha thumbnails för de scrapade bilderna, även denna lösning förkastades då vi behöver en stor bild för att en annons skall se komplett ut på vår applikation. Diskussioner ledde till slut fram till följande slutsatser:

Upphovsmannarätten till en bild tillhör alltid fotografen såvida inte speciellt avtal upprättats mellan t ex fotograf och medieföretag. I den aktuella branschen torde sådana avtal vara ytterst ovanliga. Det sannolika är att upphovsrätten tillhör annonsören eller denne närstående person och att dessa kommer att ha stor nytta av återpubliceringen.

LBS SciencePark ansåg således att det gynnar upphovsrättsinnehavaren att vi scrapar bilden. Vad det gäller att visa annonsbilden på vår applikation ansåg det var ok, eftersom bilderna tillhör annonsören och inte annonsapplikationen. LBS SciencePark kommer att undersöka detta vidare innan lanseringen av GlobalHorseShop och under tiden kommer vi arbeta under hypotesen att detta är ok.

1. För att använda bilderna behöver vi ladda ner dem från sidan som scrapeas och sedan ladda upp på vår server. Om vi hade visat bilderna från deras server skulle detta belasta deras servrar för något vi tjänar på.
2. Vi måste alltid presentera källan, oavsett om bilder scrapeas eller inte.

5.2 Hur måste vi presentera källorna?

Hur källorna skall presenteras är också en viktig fråga. Eftersom källorna presenteras i varje annons gynnar detta upphovsrättshavaren. Frågan vi ställde oss var dock om det räckte att skriva vilken applikation vi fått adressen ifrån, eller om vi var tvungna att länka direkt till originalannonsen. LBS SciencePark ansåg dock att vi borde länka direkt till originalannonsen. Detta eftersom vi då länkar information direkt till annonsören, som vi tagit material ifrån.

5.3 Vilken text och vilka bilder får vi ta från andra sidor?

När det gäller att scrapa hem bilder och text arbetar vi, utefter LBS ScienceParks direktiv, under hypotesen att det är ok. LBS SciencePark kommer sedan att undersöka detta vidare vid lanseringen då de är dom som har det juridiska ansvaret. Vad det gäller lagar i olika länder ansåg inte Science Park att vi behövde undersöka det i dagsläget, då de inte skiljer sig i de länder som är aktuella just nu, de vill säga Sverige, Norge, Danmark och Finland.

En säkerhetsventil är att GlobalHorseShop vid lanseringen kommer att betraktas som en betaversion. Det kommer att framgå att webbplaster som inte önskar ”scrapas av” kan informera om detta mailvägen. Scrapingen av den aktuella webbplats kommer då att upphöra.

I grunden tror LBS SciencePark att dessa webbplatser kommer att vara positivt inställda eftersom GlobalHorseShop starkt kommer att bidra till att driva trafik till originalannonserna.

6 Tekniska frågor

6.1 Hur fungerar scrapingtekniken?

Med hjälp av Science Parks scrapingmotor, Umbrella, hämtas information från andra applikationer som sedan presenteras på GlobalHorseShop.

Umbrellaservern går in en sida vi har valt, sparar sedan ner all information i en lång sträng. Strängen söks sedan igenom efter informationen vi vill komma åt. Sen letar vi manuellt efter mönster i denna sträng som sedan matas in i servern. Utefter dessa mönster tar servern ut den information vi vill ha och lägger in i GlobalHorseShop's databas.

6.1.1 Regular expressions (regex)

Regex används för att bestämma vad som ska hämtas från andra applikationer. Genom ange speciella parametrar kan man ta bort/hämta information eller göra både och.

Exempel på regex:

```
"<tag>Regex</tag>"
```

För att ta ut **Regex** från citatet använder vi följande sats:

```
<tag>(.*?)</tag>
```

Punkt stjärna frågetecken tar allt som står mellan parametrarna utanför så om vi hade haft något mer låt oss säga: "<tag>Regex är kul</tag>"

Då kommer vi att få ut "**Regex är kul**" genom att använda oss av samma parametrar "(.*?)".

För mer information om regex och fler parametrar se:

http://en.wikipedia.org/wiki/Regular_expression

6.2 Hur implementerar vi scrapingtekniken?

Ett instickningsprogram behövs göras för varje enskild sida vi scrapar då varje sida är uppbyggt annorlunda.

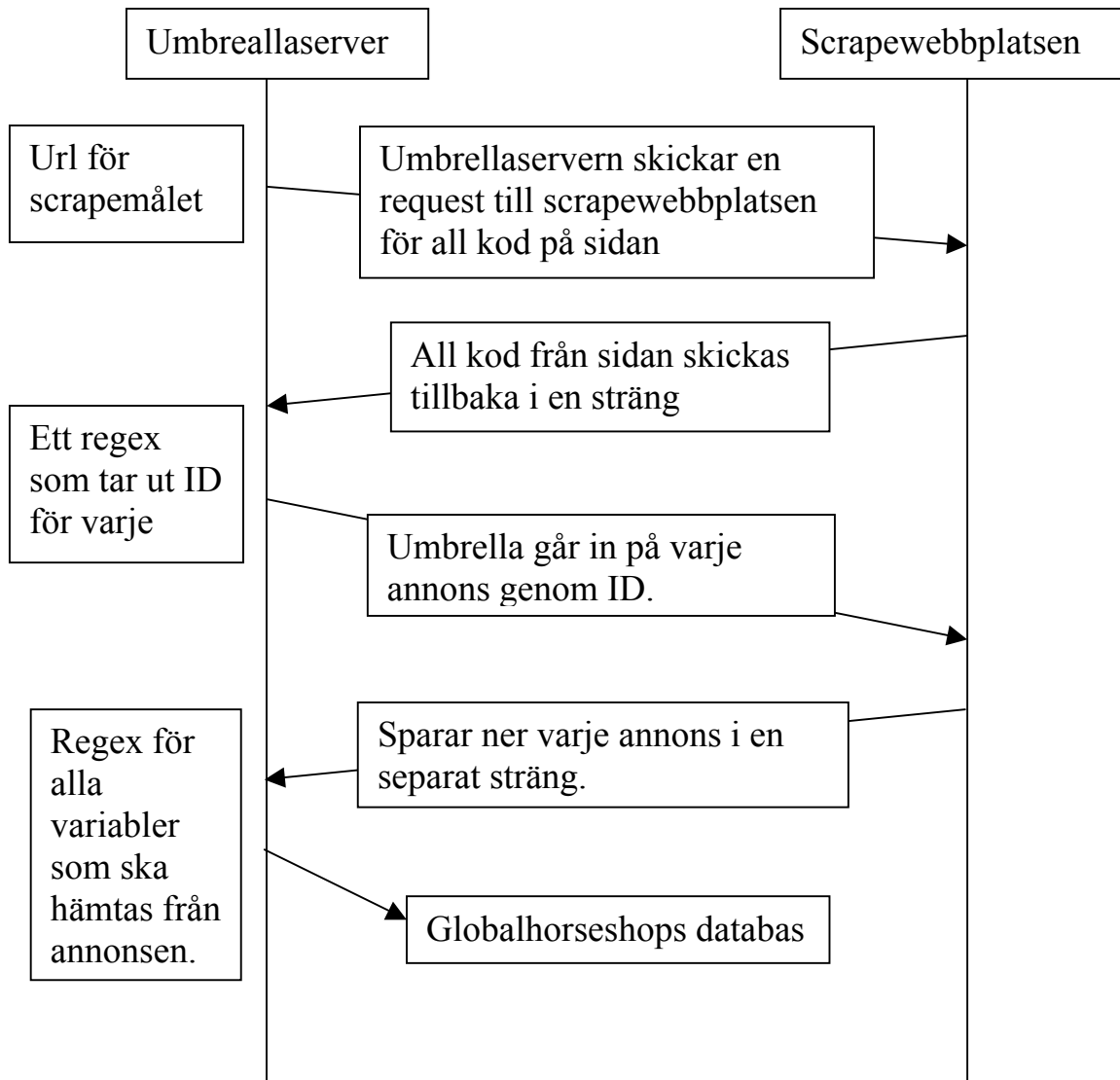
För att scrapa information från en sida skapade vi ett plugin för

Umbrellaservern. I pluginet fick vi först ange en url till där sidan listar upp sina annonser, exempelvis <http://horselink.se/hastborsen.php?sec=annonser>.

Pluginet sparade då all kod på sidan till en sträng. I denna sträng fick vi sedan manuellt hitta ett mönster till länkarna till varje annons som vi sedan lade in i Umbrellaservern. Vi var sedan tvungna att få ut varje annons ID.

Insticksprogrammet behövde sedan ett url som vi sedan lade till ID för varje annons på, för att få ut den kompletta adressen till varje annons. Väl inne i en annons letar vi efter alla enskilda variablers omkringliggande taggar. När vi sedan fått ut alla de variabler vi behöver matar vi in allt i databasen. När annonserna läggs in i databasen tilldelas dom ett unikt ID som fungerar som nyckel.

När insticksprogrammet är klart skapas en .dll fil som sedan laddas upp till Umbrellaservern.



Allt detta skapas helt fristående från gränssnittapplikationen. Detta medför att man enkelt kan lägga till instickningsprogram eller modifiera de nuvarande utan att behöva ändra i applikationen. Detta kan vara nödvändigt om det kommer nya aktörer på marknaden som är värda att scrapa eller om en applikation vi scrapar från ändrar sin design.

6.3 Hur löser vi att GlobalHorseShop ska finnas i olika språk?

Om användarens ip kommer ifrån Danmark så importeras en välkomstfil till första sidan som heter "welcome_dk.txt". All fritext finns i separata textfiler som sedan importeras till applikationen beroende på vilket land ipn kommer ifrån.

6.4 Skall vi ha två separata tabeller, en för scrapingresultat och en för annonser som läggs upp direkt på GlobalHorseShop?

Umbrellaservern lägger automatiskt in det scrapade resultatet i en Microsoft databas. Denna valde vi att ha som enda databas för vår applikation.

De scrapade resultaten läggs in i en tabell som heter ”ads”, denna har vi valt att ha som enda tabell för annonser. Vi såg ingen fördel med att ha två separata tabeller för scrapade annons och för de som var upplagda direkt på GlobalHorseShop. Vi har istället en kolumn i tabellen som visar om dom är scrapade eller inte.

Nackdelen med detta är att php är optimerat för mysql och om man använder det med Microsoft databas, kan det medföra vissa buggar, men då våra sql-satser inte är så komplexa påverka detta inte vårt projekt.

6.5 Hur ser vi till så att vi inte får två av samma annons?

De flesta annonser som läggs upp på flera sidor har sällan exakt samma text eller rubrik. Det enda som brukar vara likadant är bilden, och att göra ett script som söker efter bilders likhet hade medfört en hög serverbelastning. Vi har även valt att inte implementera detta då ett av våra delmål med detta projekt är att användarna till slut skall lägga upp alla sina annonser direkt på GlobalHorseShop, om detta mål uppnås kommer vi inte att ha problem med dubbla annonser.

6.6 Hur länge skall annonser få vara kvar på GlobalHorseShop?

Vi tar inte bort annonserna utan vi lägger dem i ett arkiv. Detta för att användare skall kunna få en uppfattning om hur priserna för en viss häst låg för exempelvis ett år sen, om så önskas. Vi har valt att ha en tidsbegränsning på en månad på annonser. Det kommer finnas ett alternativ i sökfunktionen där användaren bestämmer om sökningen även skall söka i arkivet.

6.7 Hur ofta uppdaterar vi informationen vi tar från andra databaser?

Vi scrapar varje webbplats individuellt med 5 minuters intervall eftersom när det laddas upp som mest annonser på de webbplatser vi scrapar så kommer det ungefär en annons var 5e minut.

6.8 På vilket sätt ska annonsören kunna ladda upp bilder?

För att hantera bilduppladdningen skrev vi ett phpscript som skapar 3 olika bilder av den bild som användaren laddar upp. En liten bild för thumbnails (50x50 pixlar), en mellanstor bild som syns på annonserna (255x255 pixlar) och en som är original storleken. Dessa bilder laddades sedan upp i en mapp som döptes efter annonsens id. Detta för att vi inte skulle behöva downsampla bilderna.

6.9 Hur löser vi bildscrapingen?

När vi scrapade en annons som hade en bild, lades bildens url in i databasen. Vi skrev sedan ett script som kördes var 5e minut som sökte igenom databasen, sökte upp vilka annonser som hade ett url i picture kolumnen. När scriptet fick en träff i databasen laddades bilden ner och skapade 3 bilder, en liten bild för thumbnails (50x50 pixlar), en mellanstor bild som syns på annonserna (255x255 pixlar) och en som är original storleken.

6.10 Vilka programmeringsspråk skall vi använda oss av?

Språk som vi valt att jobba med är php, javascript, xhtml, css och C#. Scrapingmotorn är skriven i C#, ett objektorienterat programmeringsspråk som används i utveckling av .NET applikationer. Vi använder det för att skapa plugins till scrapingmotorn som hämtar hem annonser från andra webbplatser.

6.10.1 Val av språk

I detta projekt har vi valt att använda oss av php, dels eftersom som vi redan har en baskunskap i detta och dels för att det är gratis att publicera kommersiella sidor i detta språk. Php är ett serverbaserat språk med andra ord det körs bara på servern. Språk som även används är xhtml, css och javascript som körs på klientens dator.

En annan anledning till varför vi valde att använda oss av php är på grund utav att SciencePark övergår mer och mer till php i sina projekt, då detta är gratis.

7 Innehållsfrågor

7.1 Vilket material ska vi utgå ifrån när vi formar gränssnittsapplikationen?

7.1.1 Målgrupp

Målgruppen är indelad i en primär och sekundär grupp.

Den primära gruppen är personer som vill sälja eller köpa sin häst.

Den sekundära gruppen är personer som inte direkt är intresserade av att köpa eller sälja, utan bara är intresserade av branschen, som istället besöker en webbplats för att få en uppfattning om marknaden.

7.1.2 Undersökning

Vi utformade ett formulär i samarbete med LBS SciencePark med frågor som skulle hjälpa oss med vad som är viktigt för en möjlig köpare eller säljare av en häst.

LBS SciencePark tillhandhöll en lista med mailadresser till relevanta personer i branschen. Formuläret skapades för Internet och vi skickade sedan ut en länk med ett mail där det stod vad syftet var med frågorna.

Målet med denna undersökning var att ha kvalitativ undersökning istället för kvantitativ.

Vi fick en svarsfrekvens på 42.5%. Vi fick även många personliga mail där frågorna var svarade med en mer personlig åsikt.

7.1.3 Frågorna

Frågorna redovisas i bilaga 1.

7.1.4 Resultat

För att få underlag för formuläret frågade vi om svarsdeltagarna hade sålt eller köpt en häst, med detta såg vi att alla som har köpt en häst har även sålt en häst. Detta medför att attraktionen till sidan ökar för dem som sålt en häst eftersom det resulterar ofta till ett nytt köp.

Vid köp och sälj häst var inte hästtypen intressant för oss utan en fråga för att göra det lättare för svarsdeltagarna att relatera till frågorna efter.

30 % av dem som köpt en häst har gjort det genom Internet, Här har vi en chans att bidra till den nivå genom att göra en attraktiv och lätt översiktlig sida där användaren som säljer en häst även ska kunna köpa en häst. Eftersom vår statistik endast utgår från köparens perspektiv, kan vi inte riktigt se vad som sker om man säljer en häst.

En viktig aspekt som vi såg genom formuläret är att nästan alla av de frågade ville se hur hästen presterade efter att den sålts, detta bidra till en återkomst till applikationen där användaren skulle kunna se hur hästen presterat. Detta löser vi genom att annonsören ska kunna lägga upp tävlingsresultat för hästen

Vem svarardeltagarna lyssnar mest på vid ett köp av en häst, gav en delad uppfattning mellan tränare och eget omdöme, vi fick däremot lite kritik på formuläret eftersom eget omdöme spelar alltid en viktig roll inför ett eventuellt köp. Svaren borde ha omformulerats så eget omdöme hade tagits bort, andra alternativ tillkommit och eventuellt en textruta för ett eget svar.

Många vill vara uppdaterade om prissituationen av hästar, detta innebär att många användare kommer att återbesöka sidan. Dessa användare är den sekundära målgruppen.

Vad syftet var med hästköpet ger oss en överblick på information som skulle kunna visas på första sidans annonser.

Vid frågan om användarna kan tänka sig att köpa häst utomlands är inte viktigt för oss i detta skede av applikationen, men våra uppdragsgivare tyckte det var en intressant aspekt att ta med, eftersom de planerar att lansera applikationen vid ett senare tillfälle i fler länder.

7.2 Hur bestämmer vi gränssnittsapplikationen identitet?

7.2.1 Brus

Analysering av andra hästrelaterade sidor:

Hippson

Bild 1: Rött representerar reklam, blått nyheter, svart hästannonsering.

Hippson har annonsering som en biprodukt och fokuseringen ligger på reklam och nyheter. Till stor del är reklamen bruset eftersom man inte har ett gemensamt sätt att visa den på, detta gör att det är svårt att fokusera på vad sidan vill förmedla. Även om det till synes finns en tydlig navigering så är koncentrationen för stor på runtomkring informationen på webbplatsen.

Horselink:

The screenshot shows the Horselink website interface. At the top, there's a header with the 'HORSE LINK.se' logo and a 'horseCam' banner. Below the header is a navigation bar with tabs: Förstasidan, Hästannonser, Köp & Sälj, Branschguiden, Forum, and Läsvärt. A status bar indicates 'Just nu: 1263 annonser, 21083 medlemmar, 4178 foruminlägg, 1749 företag'. The main content area is divided into several sections: 'Hästannonser' (Horse Advertisements) with sub-sections for 'Hästar, utrustning och tillbehör - köpes och säljes' (Horses, equipment and accessories - bought and sold) and 'Hästannonser' (Horse advertisements). The 'Köp & Sälj' (Buy & Sell) section includes links for 'Foderhäst', 'Medryttare', 'Tillridning', 'Stallplats/bete', 'Hästjobb', and 'Hömarknad (ny)'. The 'Branschguiden' (Industry Guide) section features a search bar and a list of 'Utvällda företag' (Selected companies). The 'Forum' section has a 'Veckans annons' (Weekly advertisement) and a 'Senaste annonserna' (Latest advertisements) section. The 'Läsvärt' (Worth reading) section includes a 'Google' search bar and a 'Google-annonser' (Google advertisements) section. The right sidebar contains an 'Annonser' (Advertisements) section with a 'BioLight' ad, a 'horseCam' ad, and a 'Lässugen?' (Ready to read?) ad. The bottom of the page features a 'Hästar säljes' (Horses for sale) section with a list of 'Användbara verktyg inför hästköpet' (Useful tools for horse purchase).

Bild 2: Rött representerar reklam, blått nyheter, svart hästannonsering.

Horselink har en tydligare uppbyggnad än hippson. Reklamen tar inte lika stor plats men är ändå irriterande eftersom det är svårt att ta blicken ifrån den när den befinner sig på hela sidan, till höger, till vänster, i mitten och i headern. Det finns inte någon tydlig uppbyggnad hur annonserna är positionerade till varandra, detta gör det svårt att få en överblick på webbplatsen.

7.2.2 Färgschema

En logotyp arbetades fram med hjälp av LBS Science Park och genom den har vi sedan skapat ett färgschema. Det består av följande grundfärger blå, grå, vit, svart och röd.



7.2.3 Visuellt design

Våra målsättningar att ha en väldigt lättöverskådlig design som attraherar besökaren men även ger intrycket att det är lätt överskådligt.

7.2.3.1 Design




Headern av GlobalHorseShop består av en bild som LBS SciencePark har tagit fram, en meny och en statusbar där information som hur mycket annonser GlobalHorseShop har visas, mer information kommer att presenteras vid vidareutveckling, exempelvis hur många affärer som har genomförts.

The image shows a search bar labeled "Annonssök". It has a text input field for "Sökord:" and a "Sök" button with a magnifying glass icon.

Alla knappar och block rutor har ikoner för varje funktion, detta för att stärka GlobalHorseShops identitet.

The image shows a user registration and login form. It is divided into two main sections: "SÄLJA HÄST?" (Sell Horse) and "KÖPA HÄST?" (Buy Horse). Under "SÄLJA HÄST?", there are input fields for "E-mail:" and "Lösenord:". Under "KÖPA HÄST?", there are two buttons: "Bli medlem" (Become a member) and "Logga in" (Log in). The buttons have icons representing a horse and a key respectively.


Med loginrutan vill vi få användaren att se fördelarna för att köpa och sälja häst, därför står sälja och köpa häst som en fråga. När man för muspekaren över dessa kommer en mouseover ruta som ger användaren information för ändamålet.



GLOBAL HORSE SHOP
 universal monitor for
 sellers and buyers.

[Hem](#)
[Sök annonser](#)
[Sök användare](#)
[Köparskolan](#)
[Nyhetsbevakaren](#)
[Fråga Experten](#)
[Konto](#)
[Om företaget](#)

Just nu på GHS: 9619st annonser, 2st medlemmar



VACKERT 4-GÅNGSSTO
 Namn: Arsöl från Mariebol
 Kategori: Häst
 Disciplin: Allround
 Pris: 69.000kr
 Ras: Sto

ARSÖL är ett mycket snyggt sto med fint temperament och stabila, taktfasta gångarter. Hon är ett säker och välriden häst, okomplicerad och kan ridas av alla. ARSÖL har en stor tölt, flott trav och en rund galopp. Skritten är vägvinnande och det är lätt att växla mellan gångarterna. Hon rör sig alltid mjukt och avslappnat. Hon är tävlat och regelbundet tränad för instruktör. ARSÖL passar både som tävlingshäst, fin ridhäst i skog och mark, gå på kurser eller till avel. Har haft två fina föl.

Globalhorseshop

Hej och välkommen till GlobalHorseShop, en nystartad övervakningstjänst med inriktning på hästaffärer världen över. Här samlas annonser från de största hästannonssidorna i Norden. Detta för att samla en i nuläget väldigt utspridd marknad på ett och samma ställe.

Sök annonser

Sökord: [Sök](#)

SÄLJA HÄST?

KÖPA HÄST?

E-mail: [Bli medlem](#)

Lösenord: [Logga in](#)

Senaste annonser

Rubrik: VACKERT 4-GÅNGSSTO Region: Mellan Sverige Pris: 69.000kr	Övrigt: ARSÖL är ett mycket snyggt sto med fint temperament och stabila, taktfasta gångarter. Hon ...
Rubrik: Shetlandssto Region: Mellan Sverige Pris: 9.000kr	Övrigt: Pärilan är en mycket söt och nyfiken shetlandsponny, hanterad av vuxen. Riden och tömkörd. ...
Rubrik: perfekt allround kille Region: Södra Sverige Pris: 14.000kr	Övrigt: Pga tråkiga omständigheter så är jag nu i akut behov av att hitta ett nytt hem till min un...
Rubrik: Dressyrvalack Region: Stockholm Pris: 120.000kr	Övrigt: Ett mkt snyggt, stor, högställd och pampig valack. Född :98 e. Warsteriner- Ganesco. Tävlad...
Rubrik: 2-års sto Region: Uppsala Pris: 65.000kr	Övrigt: Brunsvart sto med stjärn, efter Akribori/Cheval Sansibar/Happy Day. van att stå uppbunden,...
Rubrik: Hingst 9 mån Region: Södra Sverige Pris: 5.000kr	Övrigt: En busig och trevlig kille som är mkt hanterad. född så sent som i slutet på september. pa...
Rubrik: Snäll och kelen tjej Region: ... Pris: ...	Övrigt: Zantana är ca 155cm i mkh lite ...

Globalhorseshop, är uppbyggt med block som är igenkänt genom en bakgrundsfärg som alltid är den samma. Alla block har en speciell funktion så som senaste annonser, login, sök.

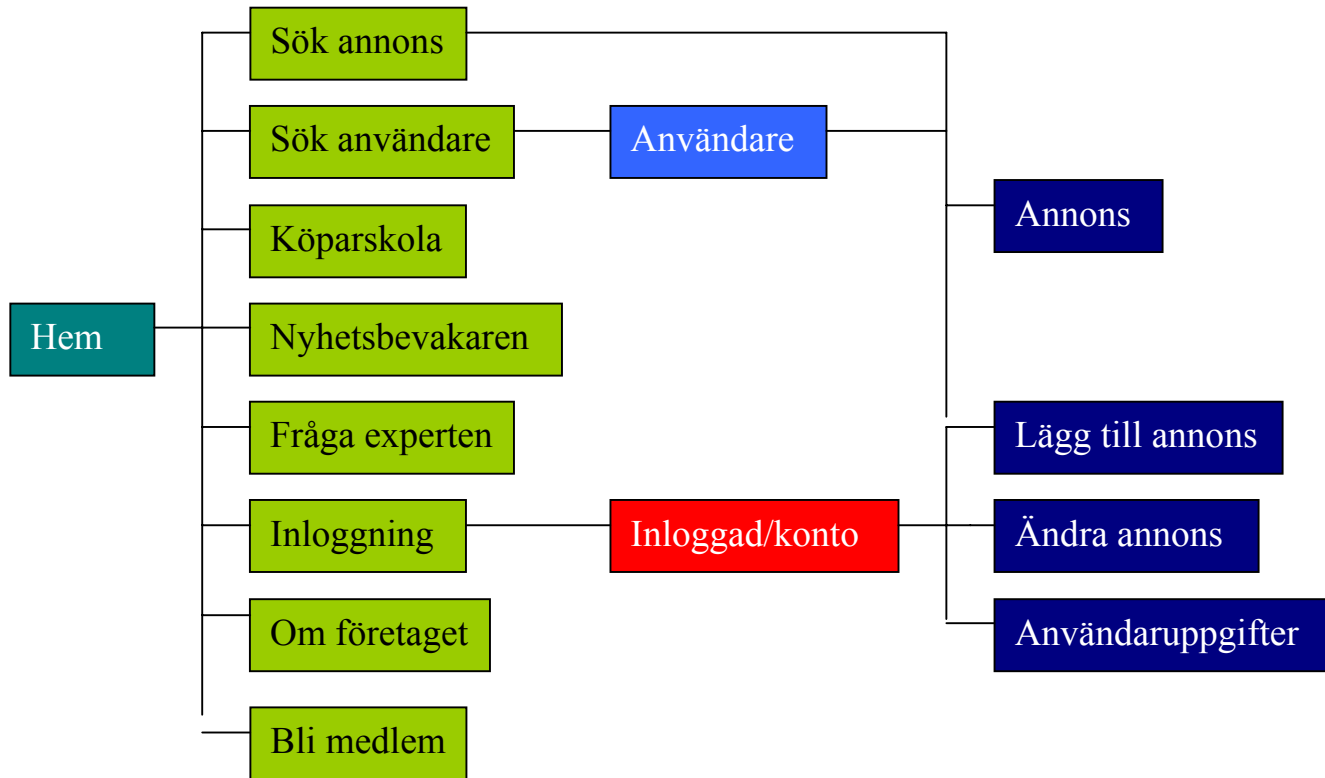
Saker som saknades på andra sidor var en sökfunktion på första sidan, en tydlig uppdelning vad som var annonser och inte, reklamen på ett och samma ställe.

Vi har gjort en tydlig uppdelning i form av block och positionering av materialet i applikationen, även en enkel sökfunktion har lagts till på första sidan. Det syns tydligt vad varje block innehåller och vad syftet är med dem.

Hur skall strukturen i gränssnittsapplikationen vara uppbyggd?

7.2.4 Navigation

Globalhorseshop navigeras genom en vågrät meny, alla sidor, annonser och funktioner så som sök kan nås genom bara ett fåtal klick, detta för att hålla oss till den enkla designen.

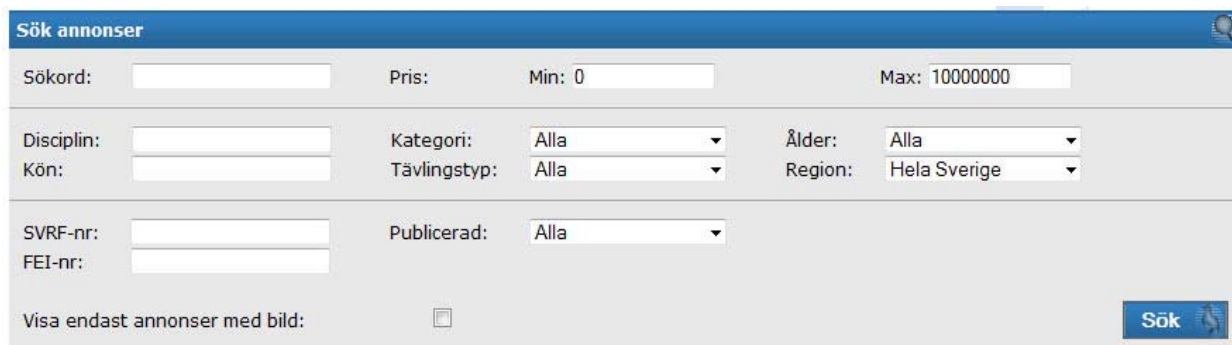


7.2.5 Funktionalitet

De flesta webbplatser med hästannonsering har väldigt mycket information, med GlobalHorseShop försöker vi istället ta bort all onödig information och hålla oss till en så minimalistisk design för att användaren lätt ska kunna integrera med webbplatsen. Det går att nå alla sidor i den vågräta menyn under headern, och för att logga in eller bli medlem behöver användaren endast klicka på en knapp.

7.2.6 Grundfunktionalitet

7.2.6.1 Annonssök



Sökparametrarna är indelade i tre olika grupper den översta är väldigt grundläggande sök, den andra är lite djupare sökparametrar och den sista är om man vill söka på en speciell häst.

På första sidan valde vi att ha en enkel sökfunktion, där användaren kan söka på valfritt ord. Resultatet av sökningen presenteras under annonsökningen oberoende på om sökningen kommer från första sidan eller från annonsök.

7.2.6.2 Användarsök

Användarsök implementerades för att göra det lätt för användare i applikationen att snabbt få en överblick på alla annonser som en specifik annonsör har, eller se en användares information så som en hemsida eller kontaktuppgifter.

7.2.7 Mervärdes funktioner

7.2.7.1 Medlemskap

För att ladda upp annonser på applikationen, eller ställa frågor till en av experterna, måste man vara medlem. Användaren behöver bara klicka på "bli medlem", sen fylla i några enstaka uppgifter och sedan klicka på skicka. Därefter är användaren redo att ladda upp annonser eller ställa frågor till experten. Målet var att det skulle vara enkelt och få medlemskap så snabbt som möjligt.

7.2.7.2 Bli medlem

En kort registrering finns där vi valt att använda e-post som inloggningsnamn, eftersom den är unik och ingen kan ha samma e-post som exempelvis ett namn. En annan viktig aspekt här är att vi valt att man kan lägga till sin hemsida. Detta beror mycket på att det finns många stall som registrerar sig och det är en stor fördel för dem att kunna lägga till hemsidan, eftersom det går att gå in på en specifik användare och leta efter information. När man klickar på användaravtalet så kommer en ruta upp istället för fördelar som köpare/säljare.

Argumenten för att bli medlem som antingen köpare eller säljare valde vi att presentera som en lista för att få ut effekten av att det är en fördel att registrerar sig på sidan.

7.2.7.3 Nyhetsbevakaren

Nyhetsbevakaren är en applikation som LBS Science Park skapat som hämtar hästrelaterade nyheter. De implementerades som en del av applikationen.

7.2.8 Begränsningar

En begränsning är att det inte går att hämta information som inte finns från andra annonser vilket gör att annonser som inte är gjorda på GlobalHorseShop kan ha mindre information. Vi försöker att göra det enklare för användaren, genom att ha en direkt länk till annonsens ursprung, där de kan se vilka parametrar som finns. Annonsinformationen anpassas dessutom beroende på mängden information från den ursprungliga annonsen.

7.3 Hur presenterar vi bäst informationen för användaren?

Sidorna vi tagit upp visar att det var för mycket fokus på bruset, så som reklam, nyheter eller Community. Vi bestämde oss därför för att göra en så minimalistisk applikation med bara det mest väsentliga i fokus, så som annonseringen av just annonser. Reklamen på Globalhorseshop kommer att ligga till höger om innehållet i en kolumn.

8 Utförande/process

8.1 Metodik

Vi har valt att delvis använda oss av en modell baserad på SCRUM modellen. SCRUM är en metodik för systemutveckling. Metodiken innebär att man arbetar med nya delmål varje ”sprint”. En sprint är en tidsperiod som kan sträcka sig från en dag till en månad. Under varje sprint är man till 100 % fokuserat på detta delmål. Efter varje sprint utvärderas och testas resultatet. I vårt fall har vi valt att jobba med sprints på en arbetsvecka, där vi börjar på måndag och på fredagen går igenom vårt resultat och bestämmer oss för nästa sprint.

8.2 Introduktion till Globalhorseshop

Efter att LBS Science Park introducera GlobalHorseShop för oss, skaffade vi information om hästbranschen. Grunden för att genomföra projektet fick vi genom LBS Science Park, enkäten och besöket vi gjorde på ponny SM.

8.3 Framtagning av material

Vi besökte diverse applikationer för hästar för att få en överblick på hur de var uppbyggda och hur informationen presenterades. Vi hade en diskussion om vilka funktioner som var viktiga och hur de skulle kunna tas fram till vår sida. LBS Science Park gjorde en undersökning av de viktigaste applikationerna inom hästbranschen. Vi fick en lista med alla applikationer som hittades, vi gick igenom och rangordnade dessa efter hur mycket annonser varje sida hade. Applikationerna valdes ut beroende av hur många annonser vi ansåg sidorna skulle generera per dag. Efter det scrapade vi sidorna genom att skriva plugins till Umbrella, som står beskrivet i avsnittet om scraping på tekniska frågor.

8.3.1 Vår första anblick på GlobalHorseShop

Registrering ska krävas så inte vem som helst kan lägga upp annonser som inte innehåller hästar, på det sättet går det att spåra vem som lagt upp annonsen genom användaren istället för att gå igenom stora logar för webbplatsen. Detta följs upp genom att ge användare möjlighet att kunna registrera.

En aspekt vi såg redan från början var skillnaderna mellan scrapade annonser och våra egna annonser. Skillnaderna var stora, detta löstes genom att anpassa den enskilde annonsen beroende på information som scrapades.

Vi tog upp om man skulle kunna lägga upp hästar som ej var för annonsering, exempel för eventuellt släktträd eller bara för att visa upp sin häst. Genom diskussioner med LBS SciencePark luta det för mycket mot communityhållet och det var inte acceptabelt för projektet. Till stor del för att vi inte ville göra en nytt Lunarstorm för hästintresserade, utan Globalhorseshop ska kretsa kring annonseringen.

Angående scrapingen kom vi fram till en del frågor:

Webbplatser som endast har kontakt med annonsören med ett formulär? En lösning som vi har är att om det inte finns något telefonnummer får användaren gå in på originalannonsen genom länken till källan vi har för varje annons.

Hitta gemensamma mönster för scrapade annonser? Detta gick inte, utan varje enskild webbplats måste ha ett eget mönster eftersom alla webbplatser är uppbyggda olika.

8.4 Utveckling av applikationen

För att kunna börja utveckla GlobalHorseShop var vi tvungna att skapa ett tekniskt underlag, som hela applikationen står på. Därefter började vi arbeta på en design, som kunde fungera ihop med målen, vi satt upp för GlobalHorseShop. Vi gjorde olika design- lösningar för de funktioner som var viktigast i applikationen. Slutligen kom vi fram till en design som visades för LBS SciencePark och godkändes. I stora drag finns grunddesign kvar, men en del ändringar har fått göras. Exempelvis har den minimalistiska designen försvunnit lite på grund av önskemål från Science Park. SciencePark ville visa senaste annonser på två olika ställen på startsidan, detta för att skapa en mer levande startside.

8.4.1 Förstasidan

Vår första fokus var att skapa en översikt av de åtta senaste annonserna där det finns information om varje annons i begränsad form. Informationen vi valde att visa i de senaste annonserna är; bild, rubrik, region, priset och delar av övrig text. Vi har en inloggningsruta där det går att logga in, bli medlem och där det finns även ett kortare argument för att bli medlem. Det finns en senaste annons som visas på första sidan. Ett krav, för att en annons ska komma med, är att det finns en bild i annonsen. Vi har en liten välkomsttext som visas om man inte är inloggad annars visas en defaultsök som anpassas för varje användares vanligaste sökresultat Defaultsök kommer att vara en vidareutveckling av GlobalHorseShop. Det finns en simpel sökruta, där det går att söka efter hästar. När man söker med denna funktion, kommer man till sökfunktionen söka annonser.

8.4.2 Annonssuppbbyggnad

Vi började implementera så att användare kunde lägga upp annonser i GlobalHorseShop genom att vara medlem. Sedan gjorde vi så det gick att ändra i annonsen. För att göra det lätt för användaren, försökte vi att göra ändringen och skapandet av annonser, med samma gränssnitt.

Rubriken av annons ligger högst upp i applikationen för att vi ville göra den så tydlig som möjligt. Tidpunkten när annonsen publiceras, finns strax under rubriken.

Bilder i annonser kom mycket senare än själva annonsens uppbyggnad, eftersom det kräver en del php-funktioner, för att det ska fungera. För våra egna användare av vår applikation, finns det möjlighet att ladda upp max fyra bilder. Denna egenskap fungerar enbart för annonser skapade på GloabHorseShop.

Vi skapade en dynamisk informationsruta på annonserna, eftersom scrapade annonser inte alltid har lika mycket information som vanliga annonser. Vad som händer är att informationrutans storlek anpassas efter hur mycket information det finns i den specifika annonsen.

En juridisk slutsats som vi var tvungna att ta hänsyn till var att scrapade annonser måste ha en källhänvisning. Detta löste vi genom att lägga till källan längst ner på sidan.

8.4.3 Scraping

Gemensamt för alla annonssidor kom vi fram till att annonserna skall ha rubrik, pris, text, bild (dock kan mängden bilder variera), kontakt och region. Vissa sidor ger möjligheten att kontakta annonsören genom ett formulär på sida, här uppstår ett problem eftersom det inte går att scrapa ett formulär. Lösningen blev att vi tog telefonnummer istället för e-post till annonsörerna från scrapade annonser. Under kontaktinformationen finns alltid en länk till originalsidan (källan) på scrapade annonser. Om köparen vill kontakta annonsören genom e-post går det genom att följa länken till originalannonsen.

Andra intressanta attribut var; kön, ras, ålder, mankhöjd, färg, föräldrar. Dessa fanns dock inte på alla webbplatser vi scrapade ifrån.

Detta krävde en dynamisk lösning på allt scrapat material, som visas på vår sida. Genom att ta fram mönster i andras applikationer kunde vi bit för bit ta ut hela annonser.

9 Testning

Under projektets gång har vi successivt testat programkoden som skrivits under projektet. Vid större funktioner har en mer ingående testning gjorts för att försäkra att andra delar av applikationen fortfarande fungerar. I sluttampen av projektet gjorde vi användartester som beskrivs i form av scenarion nedan.

9.1 Användartest 1

För att ta del av användartestet se bilaga 3

9.1.1 Bakgrund

Testet utfördes på Campus i Helsingborg vid en dator. Personen i fråga har viss datorvana men ingen erfarenhet vad det gäller hästar.

9.1.2 Slutsatser

Vi kan dra slutsatsen att den lätta gränssnittet och den enkla designen har fått effekten för användaren och att det är lätt att integrera med applikationen. Det som behöver tittas över är menyvalet, konto och hur användare för första gången hittar hur man skapar en annons. Även en länkförtydning för användaravtalet när man blir medlem behöver ses över.

9.1.3 Åtgärder

Lägga till länkar till annonsskapandet på första sidan när man är inloggad. Göra en tydligare länk för användaravtalet.

9.2 Användartest 2

Användartestet finns i bilaga 4.

9.2.1 Bakgrund

Testet utfördes på Campus i Helsingborg vid en dator. Testpersonen var i 40-års åldern och har en bakgrund inom hästbranschen som säljare och köpare av hästar. Personen jobbar inom IT så var en van datoranvändare, hade tillgång till internet både på jobbet och hemma.

9.2.2 Slutsatser

Vårt största problem här är hur användaren förstår den avancerade sökningen och användarsökningen i menyn.

9.2.3 Åtgärder

På grund av testresultatet drog vi slutsatsen att menyalternativen behövde ändras. Vi ändrade sök till Sök annons och användare till Sök användare.

9.3 Kontinuerlig testning

Vi har kontinuerligt testat all programmeringsfunktioner efter implementation, även all designprogrammering har testats gentemot de populäraste webbläsarna. Html kodningen följer standarden av xhtml 1.0 och designkodningen följer standarden av css 2.1. Följande webbläsare är applikationen anpassad för: Firefox, Opera, Safari, Internet Explorer 7 och Internet Explorer 6.

10 Resultat och diskussion

10.1 Utvärdering av förundersökning

Förundersökningen innebar, vid många fall, att diskutera frågor med LBS SciencePark, där vi gemensamt kom fram till lösningarna.

Resultatet av förundersökningen gav oss en stadig utgångspunkt när vi skulle skapa applikationen.

De tekniska frågorna klargjorde hur arkitekturen skulle se ut. Den består av tre huvudkomponenter: Umbrellamotorn, databasen och själva gränssnittet.

Frågorna angående innehållet visade hur vi skulle utforma applikationen.

10.2 Juridik

Genom undersökning och diskussioner med LBS Science Park kom vi fram till att vi får scrapa text från annonser, utan någon restriktion men måste ge källa till originalannonserna. Använder vi texten i ett nytt sammanhang, exempel ett nytt gränssnitt, får vi rätten till texten. Bilder däremot skapar ett nytt problem eftersom upphovsrätten alltid faller på personen som tagit bilden. Vi har under projektets gång arbetat under hypotesen att det är ok att scrapa bilder och text LBS ScienceParks har det juridiska ansvaret för GlobalHorseShop och de har ansvaret att undersöka det vidare innan en eventuell lansering.

Det enda som förändrades kring de juridiska frågorna var att vi tvingades skriva ett script, som laddar ner bilderna från sidan vi scrapar ifrån och lägga dem i vår databas.

10.3 Utvärdering av applikationen

Et av målen med applikationen var att samla hästannonser från hela hästmarknaden i Skandinavien till en och samma webbplats. GlobalHorseShop får in cirka 450 annonser per dag, vilket resulterar i att redan på lanseringsdagen kommer det att finnas några tusen annonser.

Vi har lyckats att skriva plugins till Umbrellaservern som scrapar tillräckligt med information från varje annons för att det skall se komplett ut när vi presenterar den. Varje annons har en rubrik, ett pris och en övrigtext, sen scrapar vi även 2-6 olika variabler till, beroende av vilken applikation vi scrapar ifrån. Ungefär två tredjedelar av alla annonser som scrapas har bilder. Den ursprungliga tanken med en minimalistisk design har försvunnit lite efterhand eftersom LBS SciencePark ville ha mer och mer och mer information på förstasidan. Vi har fortfarande försökt att hålla informationsmängden så låg som möjligt - förstasidan innehåller dock mycket info.

Många funktioner är inte implementerade, detta på grund utav att vi har under hela projektets gång haft som högsta prioritet att göra en fungerande applikation som får in en stor mängd annonser om dagen. Då detta är huvudsyftet med GlobalHorseShop. Fråga experten och köpskolan är något som LBS SciencePark får implementera vid en eventuell vidareutveckling av GlobalHorseShop. Vi hade som högsta prioritet att skapa en applikation som får in många annonser från många olika källor.

10.4 Problem

Från början tänkte vi att arbeta uteslutande i php, varför vi valde att skapa en egen scrapingmotor i detta språk. Vi lade ner ca 40 timmar på denna motor, innan vi kom fram till att skapa en scrapingmotor i tid var inget realistiskt mål. Vi fick då även reda på att Umbrellamotorn tog ca 2 år att framställa. Vi valde då att lära oss använda den istället.

Det största problemet vi har haft med scrapingen är när vi ska scrapa ett siffervärde (int) från sidor. Annonssörer kan skriva in exempelvis 10.000 kr eller någon annan valuta. För att detta skulle fungera i vår databas var vi tvungna att omvandla alla dessa variabler från textsträngar till endast siffror, det vill säga vi var tvungna att söka igenom strängarna och hitta alla tecken som inte var siffror och ta bort dem.

10.5 Olösta problem

När det gäller scrapingen har vi varit tvungna att kompromissa antalet variabler vi scrapar från vissa webbplatser. Detta gällde framförallt numeriska (int) variabler. Dessa variabler får endast innehålla siffror, om de innehåller något annat misslyckas scrapingen. I många fall har vi kunnat ta bort de känsliga tecknena i variabeln, men när en webbplats har ett prisintervall istället för ett fast pris har vi fått ignorera prisvariabeln, vilket i samtliga fall har lett till att vi valt att inte scrapa från den webbplatsen då priset är en av de viktigaste variablerna.

Vissa sidor som använder sig av cookies och _POST variabler för sökfunktionen istället för _GET har vi inte kunnat scrapa då Umbrellamotorn inte är anpassad för att scrapa denna typ av webbplatser.

Dubbla annonser från scrapade webbplatser är ett problem som fortfarande består, en lösning på problemet är att ha en person som går igenom alla annonser som scrapas. Det är en kostsam och ineffektiv lösning med tanke på att vi får in ungefär 450 annonser per dag. Genom att annonsörer lägger upp annonser direkt på GlobalHorseShop istället kommer problemet att försvinna.

10.6 Diskussion

Vad det gäller metodiken vi har arbetat efter har den fungerat bra. Att fokusera helt på ett mål varje vecka har känts väldigt effektivt. Den enda nackdelen med detta arbetssätt är att under projektets gång har det framkommit en hel del mindre buggar. Detta försökte vi lösa genom att ha en sprint där vi endast löser de buggar vi har hittat under projektets gång, dock så kvarstår det fortfarande ett antal buggar som vi inte har löst.

Under tidigt stadie av projektet diskuterades om vi skulle införa släktträd för hästar. Detta skulle i så fall bidra till en nyhet vid annonseringen, genom att kunna se tillbaka i tiden för varje häst.

Frågan kom upp om släktträd och scraping skulle fungera ihop?

Vid vidare diskussioner kom vi fram till att de nuvarande sidor vi tänkte scrapa ifrån var det svårt att få ihop tillräckligt med information. Eftersom många av sidorna inte ens har med fader och moder var det svårt att få ihop information till ett släktträd. Ett annat förslag kom då upp om att använda Sveriges Ridsportsförbund databas, där alla hästar registreras. Samarbete med dem var inte aktuellt i början av projektet, vi tog beslutet att vänta med denna funktion och ha det som en möjlig vidareutveckling istället. Databasen är däremot anpassad för att det ska kunna fungera vid en möjlig vidareutveckling.

Säkerhetsfunktioner för att förhindra att oseriösa annonser laddas upp är inte implementerade. Vi har inte kommit fram till en optimal lösning för detta problem. Diskussioner har dock kommit fram till tre möjliga alternativ. Det ena är att man har någon anställd som dagligen söker igenom annonser, detta är dock ett dyrt och ineffektivt alternativ. Det andra alternativet är att man försöker skapa något slags filter som automatiskt kollar annonserna och tar bort de som verkar oseriösa, detta är ett opålitligt alternativ då det är väldigt svårt att skapa ett filter som hittar alla känsliga annonser. Det tredje och bästa alternativet i vår mening är att man på varje annons har en knapp som där användaren kan anmäla annonsen för att vara oseriös eller liknande, ett mail skickas då till en ansvarig som får avgöra om annonsen bör tas bort eller inte. Nackdelen med detta är att man behöver ha någon anställd som sköter detta men tidsmässigt är det mycket mindre krävande än alternativ nummer ett. GlobalHorseShop har en chans att slå igenom riktigt stort på marknaden då antal annonser överstiger någon annan applikation vi har sett. Vår applikation fokuserar på annonseringen och köp av hästar, till skillnad från de flesta andra på marknaden.

Det största hotet vad det gäller GlobalHorseShops framtid är de juridiska rättigheterna till de annonser som scrapas. Om någon annonsör skulle finna detta fel kan detta ha förödande konsekvenser.

10.7 Avslutning

Trots problemen vi haft under projektets gång har vi uppfyllt huvudmålet med applikationen som var att skapa en ekonomisk lönsam webbapplikation som samlar den utsprida hästmarknaden på ett och samma ställe. Även om vi inte tar hänsyn till den ekonomiska lönsamheten så är det uppenbart att om applikationen har många användare så finns det en ekonomisk vinning genom reklam.

Vi kommer att fortsätta utveckla GlobalHorseShop efter examensarbetet och kommer med stor sannolikhet att lanseras inom de närmsta månaderna.

Eftersom SciencePark vill fortsätta utveckla Ghs så betyder det att de tror på vad vi skapat vilket är väldigt positivt och vi känner att vi skapat en webbplats som kan slå sig gentemot andra.

Vad det gäller framtiden för scrapingtekniken är det högst troligen en teknik som kommer att växa framöver. Informationen på internet måste göras mer tillgängligt och finnas mer samlat. Vilka marknader som är lämpade för scrapingtekniken är dock en fråga inte enkelt att svara på.

Scrapingtekniken lämpar sig för hästmarknaden eftersom varje häst är ett unikt objekt. När det gäller exempelvis dator komponenter eller liknande hade det inte fungerat lika bra. En annan marknad som lämpar sig för scraping är börsmarknaden. Att samla alla börskurser och aktier på ett ställe, där användaren sedan själv kan anpassa vad som skall visas efter sina individuella behov.

Scrapingtekniken är en mycket användbar teknik som troligen kommer att användas mer och mer i framtiden, men det gäller att ha en förståelse av marknaden och undersöka möjligheterna innan man försöker skapa en parapyapplikation.

11 Referenser

1. Lagboken Utgiven: 2007 Upplaga: 128.
2. Bild 1: www.hippson.se
3. Bild 2: www.horselink.se

12 Bilagor

12.1 Enkät

1a. Har du köpt en häst någon gång?

Ja: 1111111111111111 (17) 100%

Nej: (0)

Om ja

1b. Vilken hästtyp?

Ponny: 111111 (6) 35%

Häst: 222222222222 (11) 65%

1c. Hur gick du tillväga?

Hästhandlare: 111111 (6) 35%

Internet: 22222 (5) 30%

Tidning: 333333 (6) 35%

2a. Har du någon gång sålt en häst?

Ja: 11111111111111 (14)

Nej: 222 (3)

Om ja

2b. Vilken hästtyp?

Ponny: 11111111 (8)

Häst: 222222 (6)

2c. Hur gick du tillväga? (fel på svar.txt kanske?)

Hästhandlare: (0)

Internet: 2 (1)

Tidning: (0)

2d. Har du följ hästens prestation efter försäljningen?

Ja: 11111111111111 (14)

Nej:

3. Om du skulle funderat på att köpa en häst, vad skulle du kolla på först?

Prestation: 11111111111111 (15)

Tidigare ägare:

Säljare:

Ålder: 44 (2)

Var hästen finns:

4. Är det viktigt att säljaren har goda "referenser"?

Ja: 111111111111 (13)

Nej: 2222 (4)

5. Vilken rådgivare lyssnar/skulle du lyssnat mest på vid ett hästköp?

Er tränare: 11111111111 (11)

Hästens ägare: 2 (1)

Eget omdöme: 33333 (5)

6a. Anser du dig uppdaterad om prissituationen av hästar?

Ja: 11111111111111 (16)

Nej: 2 (1)

Om ja

6b. Hur?

Internet: 111111111111 (13)

Tidningar: 222 (3)

7. Anser du dig uppdaterad om tävlingar och resultat?

Ja: 111111111111 (13)

Nej: 2222 (4)

8. Hur många timmar i veckan använder du dig av någon internetjänst?

0-5 timmar: 11111111 (9)

5-15 timmar: 22222 (5)

15+ timmar: 333 (3)

9. När tror du att du/ni kommer att köpa en häst eller ponny nästa gång?

6mån: 111 (3)

1år: 22222 (5)

3år: 3333333 (7)

Aldrig: 44 (2)

10. Vad är det primära syftet med köpet?

Tävling: 11111111111 (11)

Utbildning av hästen: 2222 (4)

Nöjesridning: 33 (2)

11. Ungefär vilken ålder har er idealhäst?

3år: 11 (2)

5år: 22222 (5)

7år: 33333 (5)

9år: 44444 (5)

11år:

12. Vad är rimligt att betala för den häst ni ev söker?

25kkr: 1 (1)

50kkr: 22222 (5)

100kkr: 333333 (6)

200kkr: 4444 (4)

500kkr:

800kkr: 6 (1)

1mkr:

13. Kan du tänka dig att köpa häst utomlands?

Ja: 1111111111111 (13)

Nej: 2222 (4)

12.2 Användaretest 1

12.2.1 Scenario

1. Gå in på www.egodesk.com/ghs
2. Gå in på en annons från första sidan.
3. Hitta original annonsen från annonsen om det inte är originalet.
4. Gå till första sidan av GlobalHorseshop
- 5a. Gå till Bli medlem på sidan.
- 5b. Skriv in medlems information.
- 5c. Läs användaravtalet.
- 5d. Registrera dig.
6. Logga in som medlemmen du skapade.
- 7a. Skapa en ny annons.
- 7b. Annonsen ska ha rubriken test1.
- 7c. Annonsen ska ha kategorin ponny.
- 7d. Annonsen ska ha tävlingstyp B.
- 7e. Annonsen ska ha kön sto.
8. Annonsen ska ha 2 bilder.
- 8a. Lägg till 1.jpg från c:\temp
- 8.c Lägg till 2.jpg från c:\temp
9. Skapa annonsen.
10. Gå till första sidan.
11. Gå in och ändra rubriken på den senaste annonsen du skapade till test2.
12. Gå till första sidan.

Användartestat avslutas då sista punkten är slutförd.

12.2.2 Genomförande

- 1-2. Test person hade inga problem med att komma in på sidan.
3. Test personen hittade annonserna på första sidan utan några problem. Såg längs ner på sidan att där fanns en länk och frågade om det var den, vi svarade inte men användaren testade ändå och såg att en ny sida kom upp. Första reaktionen till den nya sidan var att annonsen likna annonsen på vår sida och användaren drog slutsatsen att det var originalet.
4. Inga problem att hitta första sidan igenom genom ett tryck på headern kom användaren dit.
5. Bli medlem var inget svårt att hitta och efter att användaren tryckt dök den nya sidan upp där informationen fylldes i. Test personen förstod inte direkt att det gick att klicka på användaravtalet utan efter att tagit musen över där såg användaren hur pekaren ändrades. Tryckte sedan och texten kom fram sidan om. Tyckte att det var väldigt smart att användaravtalet kom fram istället för fördelarna som fanns innan. Tryckte sedan på knappen för att registrera sig.
6. Utan problem loggade användaren in med det nya kontot.

7-8. Användaren var inte riktigt säker på hur den skulle skapa en annons, men genom lite logiskt tänkande kom den fram till att trycka på konto i menyn och sedan se att skapa annons fanns där. Inga problem uppstod däremot när informationen om annons och bilder skulle läggas till. Användaren tyckte om att skapandet av annonsen var utseende mässigt det samma som en vanlig annons.

9. Inga problem här tydlig knapp.

10. Inga problem att hitta första sidan igen.

11. Användaren tog slutsatsen att ändring av annons också borde finnas på konto och där hittades det, inga problem vid ändringen av annonsen eftersom det är samma gräsnitt som skapandet av annons.

12. Inga problem att hitta första sidan igen.

12.3 Användartest 2

12.3.1 Scenario

1. Gå in på www.egodesk.com/ghs
2. Sök efter hästhoppning
3. Välj en av annonserna i resultatet och klicka på den.
- 4a. Gör en ny sökning med flera variabler:
- 4b. Pris min: 10000 max: 50000
- 4c. Kön: Sto
- 4d. Kategori: Häst
- 4e. Välj endast annonser med bilder.
5. Välj en av annonserna i resultatet och klicka på den.
6. Sök efter användaren test.
7. Välj användaren test i resultatet.
8. Välj en av användarens annonser.
9. Kontakta annonsören.

Användartestat avslutas då sista punkten är slutförd.

12.3.2 Genomförande

- 1-2. Inga problem att hitta sidan och söka efter hästhoppning.
3. Användaren valde en av annonser i resultatet utan några bekymmer.
4. När användaren skulle söka igen från annonsen så uppstod lite funderingar användaren var osäker på om man skulle gå till först sidan eller trycka på sök i menyn. Valet gick till sök i menyn och såg då alla alternativ som skulle skrivas in eller väljas utan några problem. Användaren var väldigt positiv till att det gick att välja att endast visa bilder och att det var lätt att se över sökningsalternativen genom de olika grupperna som fanns.
5. Inga problem uppstod här.
6. Här uppstod lite funderingar om användare i menyn om den ledde till en sökning av användare eller information om användaren som besökte sidan. Efter att ha klickat på den gav sig allt rätt lätt och sökningen gick bra.
7. Inga problem vid valet i resultatet.
8. Här sågs det snabbt var annonserna var och klickade på en av dem.
9. Kontakta annonsörer hittades lätt och ett e-post skickades iväg.