# Project Information and Knowledge Engineering

# Fase 1

# 25-11-2011

## Groep 4

Ruben Visser 4081641

Olaf Maas 4086864 

Marieke van der Tuin 4079299

Inhoudsopgave

[Inleiding 3](#_Toc309941907)

[Doel 4](#_Toc309941908)

[Functionele omschrijving van het systeem 5](#_Toc309941909)

[Product fase 1 6](#_Toc309941910)

[Data 6](#_Toc309941911)

[Implementatie 6](#_Toc309941912)

# Inleiding

In de periode van 16 t/m 25 november hebben Ruben, Olaf en Marieke samengewerkt om een eerste versie te maken van een muziek systeem, dat gebruikers muziek kan aanraden aan de hand van keuzes die vooraf worden ingevoerd.

Eerst is het globale doel voor dit systeem opgesteld. Dit doel beschrijft wat het systeem over 8 weken zou moeten kunnen. Dit zal gemaakt worden tijdens 4 periodes van elk 2 weken. Aan het einde van elke periode wordt een werkend product opgeleverd.

Vervolgens is er een lijst met functionele eisen van het programma opgesteld. Er is bepaald welke bronnen het systeem ongeveer zal gaan gebruiken, in welke programmeertalen het geschreven zal worden en hoe het systeem ongeveer zal werken.

In deze eerste fase worden de genrevoorkeuren van gebruikers gebruikt om een lijst met hoog gewaardeerde artiesten te genereren welke aan de gebruiker worden getoond.

# Doel

Er moet een systeem gebouwd worden dat mensen muziek kan aanraden die zij mogelijk leuk zullen vinden, aan de hand van de muziek die zij leuk vinden of andere opties die zij invoeren.

Hiervoor zal een gebruiker zijn profiel door kunnen geven. In dit profiel staat omschreven wat bijvoorbeeld de leeftijd en het geslacht is van de gebruiker, maar vooral wat zijn muziekvoorkeuren zijn.

Aan de hand van dit profiel en de waarderingen van andere gebruikers, zal het systeem een lijst genereren met daarin aanbevolen artiesten en nummers. Deze lijst bevat naast de naam of titel, ook links naar bijvoorbeeld YouTube of informatie over de artiest van Wikipedia.

Deze gegeneerde lijst kan vervolgens door de gebruiker beoordeeld worden. Deze beoordeling wordt gecombineerd met beoordelingen die reeds op het web te vinden zijn. Door de gebruikersinput zal het systeem steeds betere muziekkeuzes kunnen doorgeven naarmate het programma vaker wordt gebruikt.

Het systeem moet zijn informatie kunnen halen van verschillende bronnen. Deze informatie van bijvoorbeeld verschillende muziekdatabases dient door het systeem gecombineerd te worden.

# 

# Functionele omschrijving van het systeem

Via een formulier zal de gebruiker een bepaald profiel van zichzelf kunnen aanmaken, waarin bijvoorbeeld genres, favoriete nummers, leeftijd en geslacht kunnen worden aangegeven. Aan de hand van deze input zal het systeem bepaalde artiesten of nummers aanraden. Het systeem zal dit doen aan de hand van de voorkeuren van andere gebruikers, waarderingen van anderen over de muziek, genres en informatie over de luisteraars van deze muziek.

De aangeraden artiesten en nummers kunnen vervolgens door de gebruiker beoordeeld worden door middel van ‘vind ik leuk’ en ‘vind ik niet leuk’-knoppen.

Het systeem zal via het internet bereikbaar zijn. Hierdoor is het makkelijk voor iedereen te gebruiken; er hoeft niet een speciaal programma geïnstalleerd te worden. Een nadeel hiervan zou wel kunnen zijn dat het niet mogelijk is om alle muziekbestanden die op de computer van de gebruiker zijn, door te scannen.

Het systeem zal gebouwd worden in de talen HTML en PHP. Er is voor deze talen gekozen omdat deze uitermate geschikt zijn voor het gebruik via internet. Met de HTML zullen de pagina’s voor de gebruikers worden gemaakt. Met PHP wordt de achterliggende code geïmplementeerd waarmee goede suggesties kunnen worden gedaan qua muziekkeuze, welke vervolgens aan de gebruiker kunnen worden getoond. Daarnaast zal er nog gebruik worden gemaakt van een MySQL database om de gegevens over de gebruikers op te slaan.

Om de suggesties voor muziekkeuze te bepalen, zullen er diverse bronnen van het internet gebruikt worden, welke ook weer aan elkaar gekoppeld zullen worden. MusicBrainz is een van deze bronnen die gebruikt zal worden. In deze muziekdatabase is veel informatie beschikbaar over nummers, artiesten en uitgaven, met name feitelijke informatie. Verder zal LastFM gebruikt worden, een andere muzieksite waarin met name veel informatie te vinden is over welke nummers in combinatie met elkaar worden geluisterd. Naast deze muziekbronnen zal er ook gebruikt worden gemaakt van DbPedia, de voor het semantische web geschikt gemaakte versie van WikiPedia. Hieruit kan specifieke data worden verkregen over bijvoorbeeld een artiest. Met name geboortedatum, plaats, en een korte samenvatting over de artiest kunnen hieruit worden gehaald. Verder zal de YouTube API gebruikt worden voor een directe koppeling naar het aangeraden nummer.

# Product fase 1

Voor de eerste fase zal er een systeem worden gemaakt waarmee de gebruikersinput beperkt blijft tot het kiezen van favoriete genres, en de output van het systeem slechts een rijtje met artiesten teruggeeft, welke binnen het gekozen genre vallen en daarnaast een hoge waardering krijgen. Hiervoor zal gebruikt worden gemaakt van één data bron. In latere fases zal het systeem worden uitgebreid met meer input van de gebruiker, met naast artiesten ook specifieke nummers als output. Het systeem zal dan verder gebruikmaken van meerdere bronnen en deze informatie gaan combineren.

## Data

Voor de eerste fase is er gekozen om slechts van één data bron gebruik te maken. Dit is de muziekdatabase MusicBrainz. MusicBrainz is de open muziek encyclopedie waaraan gebruikers zelf data toevoegen en verbeteren. Er is allerlei feitelijke informatie in te vinden over artiesten, albums, nummers en diverse uitgaven hiervan.

Ook is de meeste muziek uit de database getagd. Deze tags zijn op de meeste punten vergelijkbaar met genres. Zo zijn er tags voor ‘classical’, ‘pop’, ‘jazz’, ‘electrical’, etc. Niet alle tags zijn even bruikbaar in ons systeem. Sommige tags zijn maar een aantal keren gebruikt of redelijk nietszeggend (bijvoorbeeld de tag ‘production music’). Hier zal dus een selectie uit moeten worden gekozen voordat deze bruikbaar is in het systeem.

Daarnaast is er informatie beschikbaar over de waardering die gebruikers geven over bepaalde muzieknummers of artiesten. Deze informatie is verre van compleet: zo zijn er slechts 160 000 waarderingen gegeven, terwijl er zo’n 600 000 artiesten in de database aanwezig zijn. De waarderingen kunnen echter wel gebruikt worden. In een latere fase kunnen de waarderingen van de muziek verbeterd worden door waarderingen uit verschillende bronnen en de gebruiker zelf, te combineren tot een goede rating.

## Implementatie

Voor de daadwerkelijke implementatie van het product, is eerst is er een keuze gemaakt over de beschikbare genres. Aan de hand van de tag-cloud van MusicBrainz (<http://musicbrainz.org/tags>), is een selectie gemaakt van tags welke vaak worden gebruikt en daarnaast ook nog een duidelijk genre aangeven. Uit deze selectie is een lijst met 10 genres gekomen waaruit de gebruikers kunnen kiezen.

Nadat de gebruikers hun keuze hebben ingevoerd, wordt er voor de meest favoriete 3 genres bijbehorende artiesten opgezocht uit de MusicBrainz database.

De data uit de MusicBrainz database wordt verkregen door middel van directe zoekopdrachten via de XML web service welke zij zelf aanbieden. Dit geeft wel als beperking dat de berekeningen van het systeem zeer afhankelijk zijn van MusicBrainz: de muziekdatabase verwerkt slechts één zoekopdracht per seconde. Daarnaast slaat MusicBrainz zoekopdrachten over, indien binnen een seconde weer een nieuwe aanroep wordt gedaan. Hierdoor gaat er veel data verloren.

Van elk van deze verkregen artiesten wordt de waardering opgevraagd via een andere zoekopdracht. De waardering is een cijfer tussen de 0 en 5. Dit cijfer wordt gewogen met de positie van het genre uit de door de gebruiker gekozen lijst.

Door het wegen van de waardering zal het meestal het geval zijn dat een artiest welke getagd is met het eerste gekozen genre, hoger eindigt dan een artiest die nummers maakt uit het 3e gekozen genre. Dit hoeft echter niet zo te zijn. Zo krijgt bijvoorbeeld een artiest uit het 1e gekozen genre met waardering 2, een totale gewogen waardering van 10. Evenzo krijgt een artiest uit het 3e gekozen genre met waardering 4 een totale gewogen waardering van 12. Hierdoor zou de (beter gewaardeerde) artiest uit het 3e gekozen genre hoger uitkomen dan de artiest die nummers maakt in het meest favoriete genre van de gebruiker.

Deze gewogen waardering wordt opgeslagen in een database. Deze database zal elke keer voor gebruik worden geleegd. Er wordt in deze fase nog geen informatie over de gebruiker opgeslagen, vandaar dat dit kan. De databasetabel van items die hieruit komt, wordt gesorteerd op aflopende grootte van waardering. Vervolgens worden hier de eerste 10 van getoond aan de gebruiker.

## Evaluatie

Deze eerste fase was met name het verkennen van enkele van de mogelijkheden die er waren om aan geschikte data te komen. Helaas zijn we er ook achter gekomen dat de huidige databron, de REST-service van MusicBrainz, verre van optimaal is: we worden sterk beperkt in de hoeveelheid gegevens die we kunnen opvragen.

De implementatie liep op zich vrij goed. Het huidige systeem doet, rekening houdend met de limitaties van de databron, wat het moet doen, en er is al code in ontwikkeling om dit systeem uit te breiden.

Wat betreft communicatie en documentatie zijn er voor de volgende fase verbeteringen nodig. We zijn iets te geneigd om snel aan het programmeren te gaan zonder dat er vooraf wordt opgeschreven wat er precies gedaan gaat worden. Dit gaat nog wel aardig met de omvang van het huidige systeem, maar zodra de omvang van de code erachter gaat toenemen zal een betere documentatie hard nodig zijn.