데이터베이스 Project Proposal

공팔포에버팀

조장 임채경 2008-11705

류길현 2008-11594

박하윤 2008-10078

유재성 2008-11662

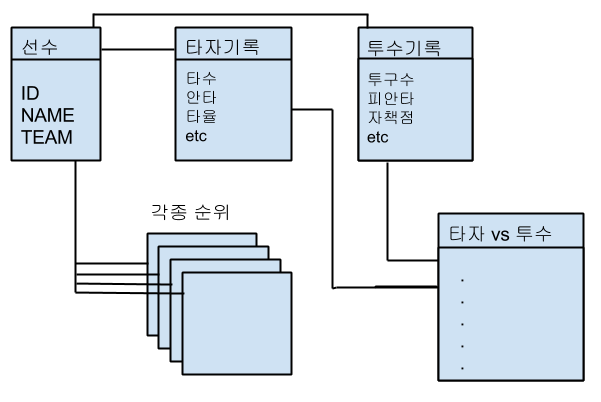
Description

1982년 시작된 프로야구는 2014년 현재 33번째 시즌을 준비하고 있다. 2008 베이징 올림픽에서 대한민국 야구 대표팀이 금메달을 딴 이후로 야구의 인기는 매년 높아지고 있으며, 그에 따라 야구 팬들의 수준 역시 높아지고 있다. 과거에는 단순히 경기를 관람하고 좋아하는 선수의 멋진 플레이를 감상하는데 그쳤다면 이제는 선수들의 세세한 기록을 찾아보고 그에 따라 승부를 예측하는 것까지 팬들의 요구 사항이 확장되고 있다. 그러한 추세에 발맞추어 우리 조에서는 KBO에서 제공하는 야구 관련 data 를 가지고 선수들의 stat 을 보관하고 user 들로 하여금 언제든 찾아볼 수 있는 야구 기록 application 을 만들고자 한다. 타율, 타점, 홈런, 방어율, 승수, 탈삼진 뿐만 아니라 좀 더 전문적인 OPS, WHIP, 피삼진, 피안타율 등의 기록들도 모두 한눈에 찾아볼 수 있는 편리한 interface 를 제공하며 우리가 보유한 data 를 통해 user 가 원하는 특정 투수와 타자의 대결을 예측해 볼 수 있는 simulation 도 제공할 것이다. 야구 팬들 중 특히 전문적인 자료를 원하는 팬들에게 유용한 application 이 될 것이다.

Uniqueness

기존에 존재하는 Android application 중에 KBO 에서 공식적으로 제공하는 application 이 있다. 이 application 은 경기 별 실시간 스코어와 팀 순위, 선수들의 기록, 관련 뉴스, 구단 정보, 하이라이트 영상, 유저 커뮤니티 등을 제공한다. 이에 반해 우리가 만들 application 은 선수들 기록에 집중한다. 네이버 스포츠, 다음 스포츠, KBO 홈페이지 등에서 구할 수 있는 일반적인 stat 들 외에 좀 더 전문적인 기록들을 제공하고, 유저 편의를 위하여 특정 선수를 DB 에서 검색할 수 있는 옵션도 제공한다. 또한, 이 application 의 가장 큰 uniqueness 는 기존에 쌓아 놓은 data 를 토대로 특정 투수와 특정 타자의 승부를 예측한다는 것이다. 2014년 프로야구 시범 경기의 data 를 정리하여 DB 에 보관한 후 유전 알고리즘과 decision tree를 사용하여 user 가 원하는 특정 투수와 특정 타자의 승부가 어떻게 될 지를 예측한다.

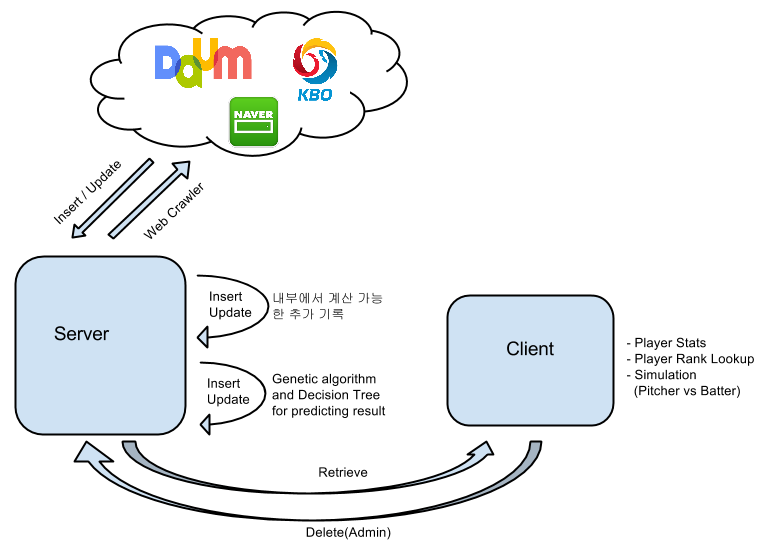
Usefulness

이 application 은 유저들로 하여금 야구 관련 상세한 기록들에 쉽게 접근할 수 있도록 하는데 목적이 있다. 실제 야구 광 팬의 입장으로서 기존의 application 은 각각의 기록에 따라 순위 내에 있는 선수들의 기록만 볼 수 있어 불편했는데 우리의 application 을 통하여 user 가 원하는 선수의 특정 기록을 한번에 검색할 수 있게 된다. 또한 네이버 스포츠, 다음 스포츠, KBO 등에서 제공하는 data 는 조금씩 차이가 있는데, 그 모든 것들을 한번에 모아서 제공하므로 다른 곳을 찾아가는 시간이 낭비되지 않는다는 것이 유용하다. 야구라는 스포츠는 팀 대 팀으로 싸우는 경기이지만 사실상 타석에서는 투수와 타자의 1대1 진검 승부이며 굉장히 확률 적인 개념이 많이 들어가 있는 스포츠이다. 우리의 application 을 통하여 선수 간 승부를 예측함으로써 경기에 대한 흥미를 높일 수 있을 것이다.

Data

KBO, Naver, Daum 등의 기록실에서 각 경기의 경기 기록 / 팀 별 타자 기록 / 팀 별 투수 기록을 기반으로 합니다.

Functionality

우리의 application 이 제공하는 기능은 크게 세가지 입니다. 타자 및 투수의 기록을 보여주는 기능, 부문별 순위에서의 선수 검색 기능, 타석에서 타자와 투수의 대결을 시뮬레이션 하는 기능입니다.

타자 및 투수의 기록은 KBO, Naver, Daum 등의 기록실에서 가져온 데이터 및 기록들을 바탕으로(Web crawler, INSERT, UPDATE), 서버에서 다른 유용한 기록들을 계산합니다(UPDATE, INSERT).

부문별 순위에서의 선수 검색 기능은 마찬가지로 타 사이트의 기록실에서 가져온 순위를 기반으로 합니다. 단, 클라이언트에서 요청한 선수 정보를 동적으로 추가합니다.

타석에서 타자와 투수의 대결은 서버에서 미리 만든 시뮬레이터를 기반으로 동작합니다. 2014년도 시범 경기 타자와 투수의 대결 데이터를 바탕으로 기계학습으로 모델을 만듭니다. CART 기반의 Decision tree 또는 유전 알고리즘 등의 기계 학습을 통해 모델을 만들고 이를 통해 타자와 투수의 가상 대결을 펼치게 합니다.

Development Plans

컴퓨터공학부 mimosa.snucse.org 웹 서버(Linux mimosa 2.6.26-2-686-bigmem #1 SMP Sat Dec 26 09:26:36 UTC 2009 i686)를 사용할 예정입니다. APM 을 이용하여 Web Application 개발할 예정이기 때문에 php, mysql, html, css, javascript 등의 언어를 사용할 수 있어야 합니다. DBMS 버전은 mysql Ver 12.22 Distrib 4.0.26, for pc-linux-gnu (i686) 입니다.

기계 학습은 python 을 사용할 예정입니다.