

BA101

BASEBALL ARTICLE WRITING

---

## Baseball Automation - Details

---

*Author:*

신승우, 김동현  
김용현, 이경수

*Typeset by:*

신승우

TEAM NoNAME

August 14, 2015

## CONTENTS

<b>0</b>	<b>Introduction</b>	<b>3</b>
<b>1</b>	<b>Source Code Details</b>	<b>3</b>
I	Crawler . . . . .	3
I.1	Article Crawler . . . . .	3
I.2	Text Broadcasting Crawler . . . . .	3
II	Simulator . . . . .	3
III	NLP handling . . . . .	3
IV	Etc . . . . .	3

## 0. INTRODUCTION

본 서류는 야구 자동작성 프로젝트에 쓰이는 소스 코드에 대한 내용을 담고 있다. 프로젝트의 모든 내용은 root 디렉토리 안에 있으며, 디렉토리는 다음과 같은 트리구조로 되어 있다.

- backup : 백업 파일. 소스 파일을 백업한다.
- documents : 관련 서류를 저장하는 폴더.
- etc : 기타 연습용 코드 등을 저장. 프로젝트가 끝난 후 지울 것.
- raw : 긁어온 raw data들을 담고 있다.
- result : 긁어온 자료를 프로세싱한 결과를 담고 있다.
- src : 소스 파일들을 저장하는 폴더.

소스 코드는 아래의 5가지로 나뉘어진다.<sup>1</sup>

- 크롤러 : 각종 사이트에서 필요한 정보를 긁어온다.
- 야구 시뮬레이터 : 야구 경기를 프로그램 내에서 시뮬레이트하여 이벤트를 추출한다.
- 자연어처리 부분 : 기사에서 필요한 정보를 추출하거나, 만들어낸 이벤트를 이용해서 기사를 만들어낸다.
- 기계학습 : 기사에 쓸 이벤트를 추출하는 classifier를 만든다.
- 기타 : 여러 정보를 처리하거나, 실행 자동화를 위한 부분을 담고 있다.

## 1. SOURCE CODE DETAILS

### I. Crawler

크롤러는 크게 2개가 있다. 야구 기사들을 긁어와서 저장하는 crawlLeagueInfo.py와 문자중계를 긁어오는 crawlTextRelay.py이다.<sup>2</sup>

#### I.1 Article Crawler

야구 기사 읽어오기.

#### I.2 Text Broadcasting Crawler

문자중계 읽어오기.

### II. Simulator

야구 시뮬레이션은 event 단위로 이루어진다. 이는 곧

### III. NLP handling

기사를 읽어오는 것과 기사를 쓰는 것으로 나누어진다.

### IV. Etc

몇 가지 기타 사항들.

---

<sup>1</sup>필요한 그림들을 그려서 넣을 필요가 있음.

<sup>2</sup>KBO에서 구질, 구속을 긁어오는 것 역시 필요하다. 아직 구현되지 않았으나 구현 필요.