TODO: insert a title here

Author:
Andrea G.B. DAMIOLI

Supervisor:
Dr. Germano BONOMI

Master Thesis

Abstract

Master degree in Computer Engineering

TODO: insert a title here

by Andrea G.B. DAMIOLI

The AEgIS Experiment at the CERN aims to verify the weak interaction principle for antimatter. This document talks about "gAnWeb", a web application designed to simplify the analysis of physical data under the AEgIS experiment. This analysis can be performed using Root Data Analysis Framework by the Linux Terminal, but a graphical interface can ensure a better user experience, ease the user training and improve the productivity. A web application is a smart way to implement the interface because allows users to avoid installations, and centralizes all the eventual modifications. This document explains the choices made during the development of this application related to the goal of the user friendly data analysis, and shows the design process that led to the final product.

Contents

Chapter 1

Introduction

First of all is important to understand at least generically what is the AEgIS experiment at the CERN and what are its goals. The acronym AEgIS stands for "Antimatter Experiment: gravity, Interferometry, Spectroscopy", this experiment aims to measure weak equivalence principle for antimatter. In the first part of this chapter are explained some particulars about this experiment, in the second part is introduced gAn Web, the main topic of this document, the application that allows the physicists to do data analysis in the AEgIS experiment environment easily, by a web interface.

1.1 AEgIS experiment

FIGURE 1.1: AEgIS's Logo

The weak equivalence principle, also known as universality of free fall, states that in the same field all bodies fall with the same acceleration, regardless of the mass and the composition. This principle has been thoroughly tested for the matter, but not for the antimatter: the most important goal of AEgIS experiment is to measure the weak equivalence principle for the anti-matter; to test this principle AEgIS measures gravitational interaction between matter (the earth) and anti-matter (anti-hydrogen). In the context of neutral antimatter, the gravitational interaction is of high interest, because it can potentially revealing new forces that violate the weak equivalence principle. Thomas Phillips, from Duke University, says: "If antimatter fell down faster, it would mean the discovery of at least one new force, probably two. If it fell up, it would mean our understanding of general relativity is incorrect". In a practical point of view AEgIS tries to mesure the time of flight and the vertical displacement of anti-hydrogen, by a moire deflectometer: this process is quite complex, and it is easier explain it by the following two images [TODO INSERT-THE-NUMBER-OF-THE-IMAGE].

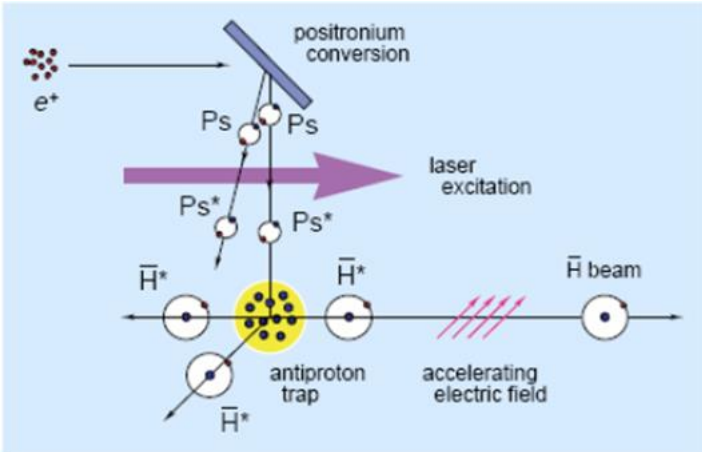

FIGURE 1.2: AEgIS's Scheme, taken from "AEgIS experiment at CERN: measuring antihydrogen free-fall in Earth's gravitational field to test WEP with antimatter" TODO INSERT-bibliographical-reference

In this first image we can see the process that allows to create some anti-hydrogen. To correctly explain this process it is better start with some definitions:

1. Positron: it is the correspondent of the electron in the antimatter. It is an anti-electron, so an electron with positive electrical charge. It is indicated by "e+".
2. Positronium: it is an unstable system consisting of an electron and a positron, bound together into an exotic atom. It is indicated with Ps.
3. Antiporoton: it is the antiparticle of the proton. Antiprotons are stable, but they are typically short-lived since any collision with a proton will cause both particles to be annihilated in a burst of energy. It is indicated with $\overline{p}$ (pronunced P-Bar).

4. Antihydrogen: it is the antimatter counterpart of hydrogen. Whereas the common hydrogen atom is composed of an electron and proton, the antihydrogen atom is made up of a positron and antiproton. It is indicated with $\overline{H}$ (pronunced H-Bar).

5. Antiproton trap: a devices that uses an axial magnetic field to transversely confine charged particles, in this case antiprotons.

The process shown in the image is the following: a beam of positrons (that comes from a 22Na radioactive source) is accelerated and driven to collide against a "positron-positronium converter" (that is a mesoporous silica film). This process creates positronium, that needs to be excited by lasers, to reach the Rydberg State. The positronium in Rydberg state is indicated by $Ps*$, it has a longer life than the unexcited positronium, and can be driven to fly into an antiproton trap.

Antiprotons are provided in this way: Protons collide with nuclei inside a metal cylinder called "target". About four proton-antiproton pairs are produced in every million collisions, and is possible to separate antiprotons from matter using magnetic fields. The following step is to guide antiprotons toward the AD (Antiproton Decelerator) where they are slowed down (it is easier work with slow antiprotons). To carry out AEgIS experiment antiproton must be trapped and conserved inside an antiproton trap, where magnetic fields force the charged antiparticles to spiral around the magnetic field lines, and electric fields confine them along the magnetic axis.

In the following step $Ps*$ and $\overline{p}$ can combine themselves to generate excited antihydrogen ($\overline{H}*$) and electrons. The antihydrogen beam is accelerated using an electric field towards a Moiré deflectometer, during the travel it decays to ground state.

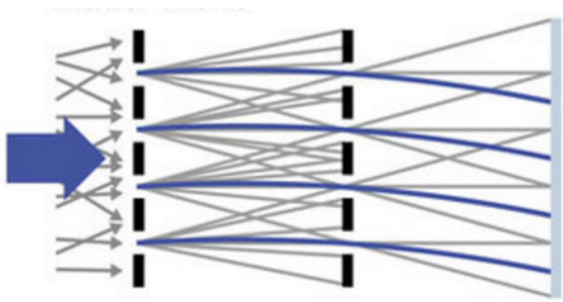

FIGURE 1.3: Moiré Deflectometer's Scheme, taken from "A http://www.nature.com/articles/ncomms5538" TODO INSERT-bibliographical-reference

In the second image is visible how does the Moirè deflectometer work. A antihydrogen beam is thrown toward two subsequent gratings that restrict the transmitted particles to well-defined trajectories. The trajectories are inflected by a force (in this case the force related to $m * g$) and follows a parabolic path. At the final part of the deflectometer there is a detector that shows where the antimatter annihilates, so is possible to compare the expected trajectories without forces with the obtained trajectories, and measure the force.

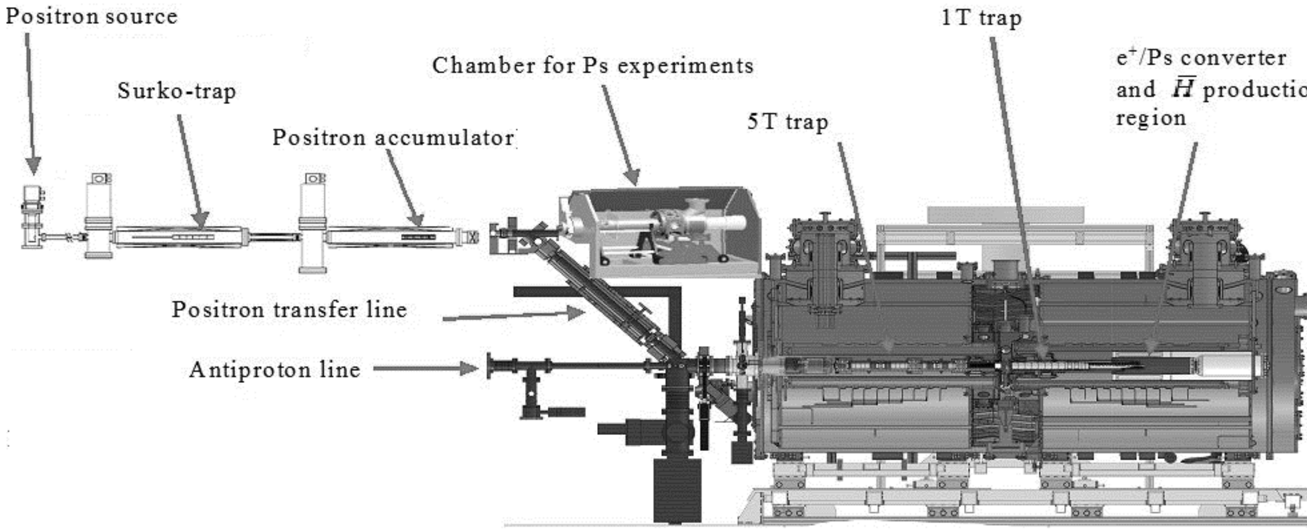

FIGURE 1.4: AEgIS apparatus set up, taken from "AEgIS experiment at CERN: measuring antihydrogen free-fall in Earth's gravitational field to test WEP with antimatter" TODO INSERT-bibliographical-reference

1.2 User friendly Data analysis: gAn Web

GAn Web is a web application, that creates a user friendly web interface, based on the most important human-machine interaction principles, between the users and a pre-existing data analysis application named gAn. This document's most important goal is explain in detail how gAn Web works, how and why it was created, what are the reasons for the choices made.

GAn Web is based on the pre-existent stand-alone program gAn, that allows users to do data analysis using a Linux terminal as interface. In turn, gAn is based on Root Data Analysis Framework, a vast and modular scientific software framework that provides all the functionalities needed to deal with big data processing, statistical analysis, visualisation and storage of physical data. GAn exploits and organizes the functionalities of Root, the resulting software is practical and achieves his goals, but a web interface can improve it in two ways:

1. gAn is a stand-alone program based on Root, installable on the user's machine; the user has to install the correct version of Root to avoid compatibility problems (Root is still not perfectly version independent: different versions can lead to different behaviours). Furthermore, this kind of program is continuously changing, the performed analysis is continuously improved, so the installed version of gAn is not final and unchangeable, and the user musts often update it. Instead, a centralized version installed on a server, with services accessible from a normal browser by the user can avoid (at least reduce) this kind of problems and be more usable.

2. a Linux terminal interface is practical for expert users, but a web based interface can be more attractive for new users, and, if well done, can be easier to use. It is important to notice that the users are physicists, not necessarily specialized in computer science, so, create a friendly and easily learnable interface can avoid them problems and time wasting.

The goal of gAn Web is to allow users to do analysis through a more friendly web interface, without install nothing on their machine. In the following image there is a schema that shows how this program is organized.

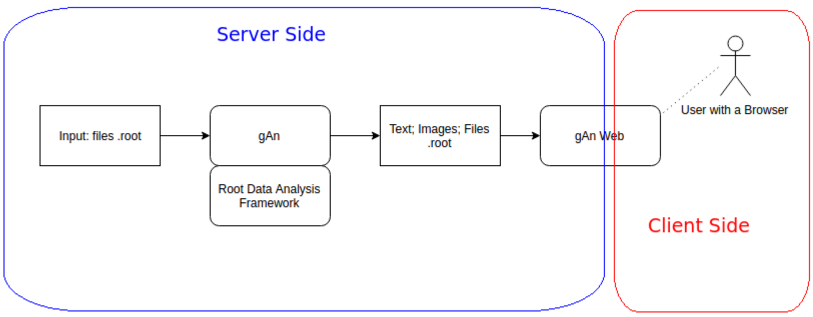

FIGURE 1.5: gAn - gAn Web simple scheme

The input of this system is represented by a set of files .root. These are raw, very big, binary files, incomprehensible to humans, generated by the hardware (mostly by detectors) of the AEgIS experiment, they need the Root Data Analysis Framework to be interpreted. These files contain a lot of information, too much and too disorganized to be helpful for the analysis. GAn can read these Root files, make an ordered and organized analysis, extract the most important information, and produce an output that consists of a text with the most important informations, comments, and eventual error logs, a folder with some images, that can summarise effectively the most important points of the analysis, and some other .root files, with useful informations that allow gAn Web to make further processing. GAn Web can close the cycle acting as intermediary between the users and gAn: gAn Web can receive requests from the users, configure gAn to satisfy these requests, and deliver to the users exactly what they need.

Chapter 2

Data Analysis

This chapter aims to explain first of all what is the "big data analysis", and why it is important in the modern world. Subsequently it will be exposed the exact use of the big data analysis technologies in AEgIS experiment, with reference to the used technologies and the choices made.

2.1 What is?

According to the John Tukey's definition data analysis is:

"Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data" (`http://projecteuclid.org/euclid.aoms/1177704711`).

The basic idea is that in the modern world almost each activity can provide a big amount of data, but only a few of them are really useful to gain interesting information. The data analysis is an structured process that allows to select the most important parts of this row data and exploit them to gain information able to answer questions, test hypotheses and approve or disprove theories. In the following image we can see the schema of this process.

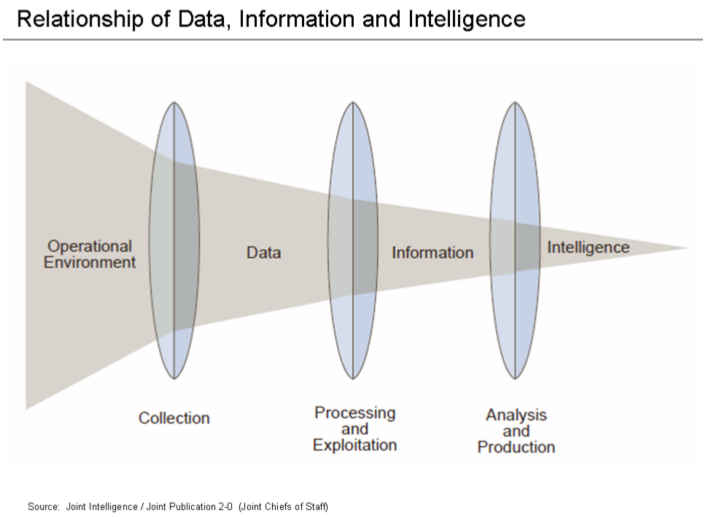

FIGURE 2.1: Here is visible a basic schema of data processes and analysis

Data analysis can be divided in some steps:

1. Data collection: data can be collected in a variety of ways. For example they may also be collected from sensors in the environment, such as satellites, recording devices, physical sensors etcetera. they may also be obtained through interviews, downloads from online sources, or reading documentation, so the analysis is feasible with a large variety of kinds of data.

2. Data processing: raw data must be well-organized for analysis: for example, placing data into row, columns, vector, etcetera.

3. Data cleaning: Once pre-processed and organized, the data may be incomplete, contain duplicates, or contain errors. Data cleaning is the process of correcting these errors, eliminating duplicates and handling incomplete data. Some ways to do this are record matching (confrontation between the records to find if there is something suspicious), validation of data (if there is the sureness that data values has to respect some limits), overall quality of existing data, de-duplication (process of removing of duplications). For particular kinds of input this process is very complex (for example vocal input needs an advanced spell-checker), for others is simple (for instance online-survey interviews made using closed choices)

4. Exploratory data analysis: in this step the data is analyzed. There are a variety of techniques referred to as exploratory data analysis to begin understanding the real content of the data.

This process may result in Descriptive statistics, such as the calculation of average or median, or in Data visualization, that allows to examine the data in graphical format, through graphics and other graphical objects.

5. Modeling and algorithms: another step is using mathematical models to find relations between different variables, such as causality or correlation. An example is the regression analysis.

6. Communication: this is the final step, and it is absolutely not trivial. Is important to find a way to report the obtained information to the user in an understandable format. The communication must be adapted to the different users, and to their requirements, to allow the data analysis to match them.

2.2 What a user can obtain from the data?

An user can benefit from the analysis of big data in various ways, in particular a working data analysis software can perform the subsequent tasks:

1. Retrieve Value: the system receives in input some cases (for instance case-A, case-B etcetera) and a set of variables (for instance variable-A, variable-B etcetera), and can shows in output the values of the variables in the set in the data cases (in case-A, case-B, variable-A = x, variable-B = y).

2. Filter: the system can show only the subset of the data that respect some conditions.

3. Find extremums, ranges and characterize distribution: the system can show the maximums o minimums values in the datasets, and in how large is the range in that the values are distributed. It is also possible to approximate with mathematical functions the statistical distributions of subsets of data

4. Find Anomalies: check the dataset to find if there are some exceptional values, that are source of interest and need an explanation (errors in the data? unknown phenomena?)

5. Clustering: the analysis is extremely easier if is possible to group different subset of data with similar characteristics. For example: in a market research, obtaining a correct clustering on the customer allows to understand the different categories of customer with different goals and needs

2.3 Relevance in the modern market

todoooo

2.4 Root - Physical Data Analysis

2.5 Data Analysis in Aegis Experiment

Chapter 3

Human-Machine Interface

labelChapter3

3.1 Human Machine Interaction principles

todo todo

3.2 Expected users

todo todo

3.3 Validation of gAn Web against HMI principles

todo todo

3.4 Modifications to match HMI principles

todo todo

Chapter 4

Used Technologies

todotodo

4.1 Web interface vs Java Fx vs Xojo

todo todo

4.2 Used Technologies and Framworks

todo general

4.2.1 PHP

todo particular

4.2.2 Javascript

todo particular

4.2.3 Bootstrap

todo particular

4.2.4 Sass

todo particular

Chapter 5

The resulting software

todotodo

5.1 A tour of the application

todotodo

5.2 Use example

todotodo

5.3 Some Screenshots

todotodo

Chapter 6

Conclusions

todotodo

6.1 Conclusions