Development of an application as a diary useful tool for bioinformatics: Biopocket

Tejedor Félez, Sergio
MU Bioinformatics and biostatistics 2018/19-2
Bioinformática estructural y diseño de aplicaciones bioinformáticas

Ortega Carrasco, Elisabeth
Prados Carrasco, Ferran

05/06/219

"If you can IMAGINE it, you can PROGRAM it"

PROJECT WORKSHEET

Title:	*Development of an application as a diary useful tool for bioinformatics: Biopocket*
Author:	*Sergio Tejedor Félez*
Consultor:	*Elisabeth Ortega Carrasco*
Delivery date (mm/yyyy):	06/2019
Certificate:	*Máster universitario en bioinformática y bioestadística*
Project area:	*Bioinformática estructural y diseño de aplicaciones bioinformáticas*
Language:	English
Keywords	*Databases, Mobile applications, WebApp*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

Ha habido un crecimiento exponencial en la generación de datos biológicos y se ha dado un aumento notable de las bases de datos y software especializados en la consulta, gestión, análisis, comparación y creación de estos datos, que suelen requerir de conocimientos específicos, y no presentan una interfaz sencilla para el usuario.

Así se ha conseguido generar mucho conocimiento, pero aparece disperso en diferentes bases de datos y no es realmente accesible para todos.

Biopocket pretende ser la solución a este problema, una aplicación multiplataforma, disponible para Android, iOS y web. Para hacerlo, esta aplicación une los diferentes software y bases de datos del campo y los presenta en una misma interfaz, que busca ser sencilla pero dotar de las herramientas principales al usuario. Entre ellas destacan el acceso a bases de datos como NCBI, UniProt o Entrez y software como BLAST o Clustal.

Es la primera aplicación del campo diseñada para ser utilizada en el móvil, respondiendo a la creciente demanda de la sociedad de poder trabajar desde cualquier lugar y tener todos los recursos en el bolsillo.

La aplicación ha sido construida con el software de Mobincube para su diseño en Android y iOS, y con javascript para su versión WebApp, alojada en el servidor Netlify. Todas las conexiones con los servidores se han establecido de forma gratuita.

La aplicación resultante es útil y sencilla a la par que potente. Es un primer paso para acercar la bioinformática a todo el mundo y en cualquier lugar.

Development of an application as a diary useful tool for bioinformatics:
Biopocket

Abstract (in English, 250 words or less):

There has been an exponential growth in the generation of biological data and there has been a notable increase in databases and software specialized in the consultation, management, analysis, comparison and creation of these data, which usually require specific knowledge, and they don't present a simple interface for the user.

This has generated a lot of knowledge, but it appears scattered in different databases and is not really accessible to everyone.

Biopocket aims to be the solution to this problem, a cross-platform application, available for Android, iOS and web. To do this, this application joins the different software and databases of the field and presents them in the same interface, which seeks to be simple but to provide the user with the main tools. These include access to databases such as NCBI, UniProt or Entrez and software such as BLAST or Clustal.

It is the first application of the field designed to be used in the mobile, responding to the growing demand of society to be able to work from anywhere and have all the resources in the pocket.

The application has been built with the Mobincube software for its design on Android and iOS, and with javascript for its WebApp version, hosted on the Netlify server. All connections to the servers have been established free of charge.

The resulting application is useful and simple as well as powerful. It is a first step to bring bioinformatics to everyone and anywhere.

INDEX

Figure list

Table list

1. Introduction

1.1. Background

Throughout the last two years I have been studying the Bioinformatic and biostatistics master's degree at the Universitat Oberta de Catalunya (UOC), and before I had been graduated in Biochemistry at the Universitat Autònoma de Barcelona (UAB).

From the Biochemistry degree I could already see the value of bioinformatics in research and its exponential growing in the field of biosciences. In my opinion, bioinformatics are one of the most powerful tools in biosciences because of many reasons, but I can highlight four of them:

- They can provide us a lot of information before and during the experiments.
- They are able to simulate some of them and predict their results.
- They are capable of comparing results through statistical analysis.
- They give us the way to share information and results.

We can compare this with any experimental case that you can imagine, because it has at least six parts:

- Observe, detect the problem and raise and hypothesis.
- Get information about it, at macro and microscopic levels.
- Perform the experiments.
- Analyze the obtained results.
- Confirm/Refuse the hypothesis, write a theory.
- Share it with other scientists.

As we can see, bioinformatics have a role in every step in research, except those related with the raising and confirmation/refuse of the hypothesis, which depend on the ability of the researcher. Of course, in spite of the fact that bioinformatics have gained value and a place in every laboratory, the use of experiments will always be necessary in research, although because of bioinformatics we will be able to reduce them, saving time, money and narrowing the use of experimental animals.

This is why I decided to study bioinformatics, and this master degree project completes my formation in bioinformatics and biostatistics at the UOC.

The project begins reviewing that in recent years there has been an exponential growth in the generation of biological data (Letondal C. et al., 2009). Both from *in vivo* biological research and from *in silico* research (Vincent K. et al., 2010). Multiple databases have appeared to store all the generated data and, in the same way, in order to consult, manage, analyze, compare those data or create new through simulations, the number of specialized software have also increased (Letondal C. et al., 2009; Bioinformatics tools and software, 2019).

So, currently exist a lot of bioinformatic platforms, software and servers that offer a wide variety of resources and information (Letondal C. et al., 2009; Bioinformatics tools and software, 2019), but if we have a look to the points I mentioned before, about why bioinformatics are one of the most powerful tools in current research, we will find out that they are clearly differentiated, separated and focused and useful for only one of them (Klopp C. et al., 2016). And, of course, each tool has its own structure, organization and requirements (Letondal C. et al., 2009; Vincent K. et al., 2010).

This makes the use of bioinformatic tools complicated, because the researchers must, not know only what they want, but where and how search or perform the analysis and if there is some previous information needed (Ahmad N., et al., 2014). This often makes bioinformatics difficult for users, because, "in which database should I search for this?" or "what do I need to search for this?" are not uncommon questions. To this we have to add the fact that the most part of the current software and platforms are not friendly for non specialized users and require specialized knowledge of each one (Letondal C. et al., 2009).

Considering all those factors we can conclude that the resources offered by bioinformatic tools are not really available for non specialist in bioinformatic researchers (Vincent K. et al., 2010). Moreover, the fact that each tool is focused on different areas of the field, makes the search or analysis harder and more complex for those researchers that use to work with them (Letondal C. et al., 2009; Vincent K. et al., 2010).

So, the problem to solve here is that there is no website or application that do a friendly integration effort for the researchers (both specialized or not), and currently, although the development of new technologies, keep being unable to have all the information in one hand (Ahmad N., et al., 2014; Althammer S. et al., 2018).

Biopocket aims to be the solution to this problem, a cross-platform application, available for Android, iOS and Web, that seeks to facilitate the day to day of a bioscience researcher, even if it is non specialized in bioinformatics.

To do this, first of all, Biopocket acts as a nexus for the different bioinformatic tools and will detail the needed information for each investigation, allowing the researcher to perform any analysis. Also, will be developed according to aesthetical factors, fitting the different resources in the same interface, which seeks to be simple and friendly but, at the same time, provide the user the main tools. By this way we delete the problematic factors commented before, making bioinformatics friendly for the user and connected all in one. These include access to databases such as NCBI, UniProt or Entrez and software such as BLAST or Clustal.

It's important to highlight that currently there is no application, neither website that do this, all in one click for the researchers, and, moreover, Biopocket is the first application of the field designed to be used in the mobile or tablet, responding to the growing demand of society to be able to work from anywhere and have all the resources in the pocket.

1.2. Project objectives

The goal of this master degree project is develop a new tool for bioinformatics with a nexus role integrating the most important bioinformatics servers and with user-friendly appearance, useful for specialized and also non specialized researchers.

To reach this goal I have defined two general objectives for the project, which can be splitted into more specific.

1.   Develop an application

    1.1.   Test and compare the different mobile/web application software developers.
    1.2.   Select the most appropriate for the app development.
    1.3.   Design and build up the application in different platforms.

2. Connect the application with the most important bioinformatic tools

    2.1. Find out and list the most important bioinformatic tools with available API code.

    2.2. Discuss with specialized and non specialized researchers the selected resources to include into the application

    2.3. Add to the application the selected tools

    2.4. Assess the obtained product.

    2.5. Identify and fix possible mistakes.

It's clear that the objectives split the project in two differentiated parts, the structural, which consists in the build up of the application, and the functional, which consists in the correct connection with other bioinformatic tools and the considerations of what a researcher need.

1.3. Materials and methods

As in the objectives, the methodology can also be splitted in two, on one hand, those methods related with the structural and on the other hand, the functional part of the application.
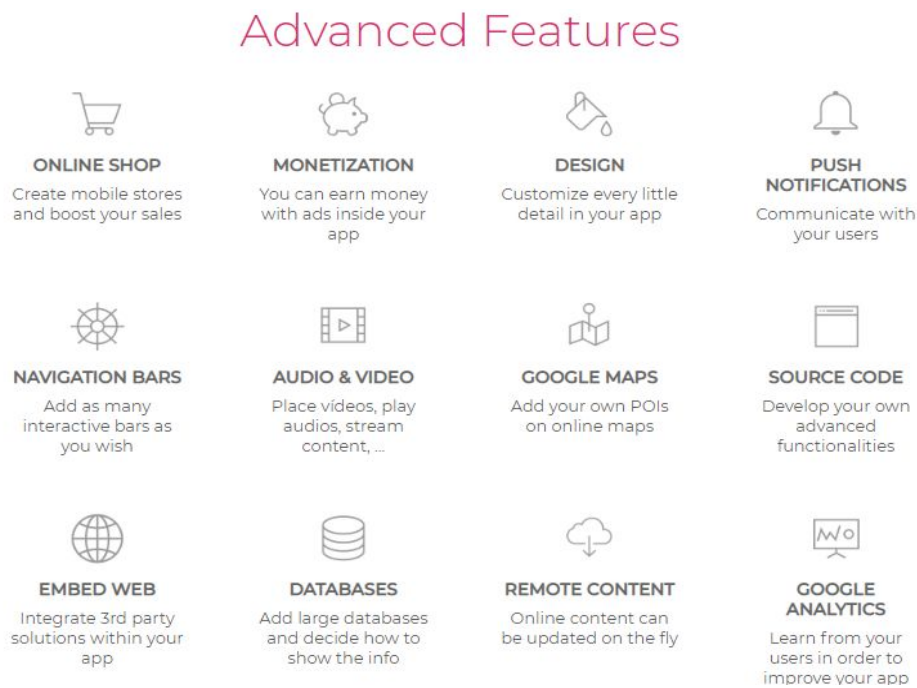
The methods used to build up the structure of the application has been changing throughout the project. It's clear that the most part of it is based on the use of an app development software, but after start working on the app development, and with the results of the surveys to scientists, I had new ideas that I must include in this project.

First, I get informed about the different mobile app developers, and decided to use the *Mobincube* platform to build the application. *Mobincube* is an online software specialized in cross-platform app building, so *Biopocket* will be available for iOS and Android (¿Qué es Mobincube?, 2019).

From the wide range of options we have when we look for an application developer, and the opportunities that the different software offer to design and build our own app, *Mobincube* stands out because is more user friendly, so I, who never has developed an app before, felt more comfortable using it. But not for that, *Mobincube* is less powerful than others. At the same time, is one of the app building platform with more options, it has a lot of advanced features (Figure 1), allowing the user to build up an app without coding, with an easy cube system and total flexibility, only have to drag, drop, copy, paste or click to make it (Funciones avanzadas, 2019).

Moreover, *Mobincube* has two more interesting features, one of them is the fact that, as I said before, the user *can* build an app without coding, but the platform let the user to introduce personal code if it's needed. In this case, it's absolutely necessary because the API code of each bioinformatic tool must be included to get connected with *Biopocket*. The other one is that *Mobincube* platform works always online, so the application will ever be available from any device at any time without neither extra software installation or data.

Advanced Features

ONLINE SHOP
Create mobile stores and boost your sales

MONETIZATION
You can earn money with ads inside your app

DESIGN
Customize every little detail in your app

PUSH NOTIFICATIONS
Communicate with your users

NAVIGATION BARS
Add as many interactive bars as you wish

AUDIO & VIDEO
Place vídeos, play audios, stream content, …

GOOGLE MAPS
Add your own POIs on online maps

SOURCE CODE
Develop your own advanced functionalities

EMBED WEB
Integrate 3rd party solutions within your app

DATABASES
Add large databases and decide how to show the info

REMOTE CONTENT
Online content can be updated on the fly

GOOGLE ANALYTICS
Learn from your users in order to improve your app

Figure 1: A list of some of the most relevant features that Mobincube offers.
From: "Funciones avanzadas", 2019 [online]

Despite this, after working a bit with Mobincube, I realized that the own code added to Mobincube will run locally if the user don't have a premium plan, so I will need an external server to carry out the connections. On the other hand, the scientist interviewed, specialized and non-specialized, point out that it would be a shame if the app was only available for Android and iOS, which means that will be only useful from the mobile. So, although one of the strong points of Biopocket is to become the first mobile app about bioscientific research, it's true that its users use to have a computer while working.

Because of this I decided to develop also a WebApp. In this way, Biopocket has become a real cross-platform application, located in an external server, and the user is able to use it from any computer or device.

So, after doing research about how to develop a WebApp I decided to use javascript language, specifically with the Node.js architecture. That's because javascript is an event based programming language, light and efficient, that allows the inclusion of API code to perform stable connections with external servers. The WebApp is deployed with Netlify and located in its servers.

The choice of *Netlify* as a server is not random. Netlify is specialized in the deployment of WebApp and works with so many languages.

Finally, the structure will be build up with *Mobincube* for Android and iOS versions of the app, linked to the WebApp deployed with *Netlify* for Desktops, in order to perform the connections with bioinformatic servers.

To this we have to add the use of *Canva* and *Adobe Illustrator* softwares for image design.

All the softwares and hardwares will be used under free licenses. This could limit the use of the *Mobincube* platform, *Netlify* server, and also some of the APIs connections.

In relation to the functional part, the methodology used is focus in the selection of the bioinformatic resources that can be included in Biopocket.

The selection took place in two steps. The first one, essential, was the search for API code in the different bioinformatic utilities. This is because, the Application Programming Interface (API) is the code that contains the language of each service, how to perform queries, how log-in, how the data is given, etc. So, it's indispensable to establish connections with the app.

After this search, I have a list of those utilities that I could include in the application. So, to know which resources would be more useful for the researchers, I decided to make a survey to some of them. It's important to remember that the objective of Biopocket is not only to integrate all the resources in one app, but also be friendly with those who don't use to work with bioinformatic programs. Because of this, I send the survey to specialized and non-specialized researchers.

1.4. Working plan

In order to carry out the project the only needed is my laptop with the required software.

As I mentioned before, there are two clear parts into this project, the structural and the functional. Of course, it does not have any sense to make only one of them, but is clear that achieve all all the functional chores but no one of the structural ones is completely illogical, because, by this way, won't be any structure to make the connections. Because of this, the first thing is develop and build up the application, and later, establish the connections.

So the tasks to achieve the planned objectives can also be splitted and agrupated in one of those parts as shown in the following table (Table 1).

Table 1: Relationship between the objectives and tasks of this project. In light blue are the structural ones and in deep blue the functionals. In each category the tasks are ordered according to chronological order.

Objective	Tasks
Test and compare the different mobile and web application software developers	- Register, and get if it's needed, in different app development softwares such as Android Studio, Xamarin (iOS), Ionic dashboard, appery.io and *Mobincube*. - Register in Heroku, Netlify, Google App Engine and Firebase - Try to make a new app with every one and see how it works - Compare the different options setting as a principal point the easiness and the range of options/features to develop the app.
Select the most appropriate for the app development	- From the comparison done before select the best platform to develop our app.

Design and build up the application	- Learn how it works - Search into the selected app building software web to deepen into its options. - Search for more information and examples on articles or internet. - Practise - Decide how many screens it would have and the function of each one and create them. - Design the content of each one: buttons, tabs... - Choose the symbols and other aesthetic factors.
Find out and list the most important bioinformatic tools with available API code	- Search into articles and asking researchers the most important (and diary useful) bioinformatic tools. - Search the internet for this applications and check if they have the API code. - Make a list with those selected.
Discuss with specialized and non specialized researchers the selected resources to include into the application	- Argue why the chosen tools are the the best to inlclude in the application. - Ask the researchers if something is wrong or they want something that is not considered yet.
Add to the application the selected tools	- Learn how API connection works and obtain the API code of the chosen resources - View other apps examples working with API code in javascript. - Link the software, databases with the different screens obtained in the structural chores.

Assess the obtained product	- Review the application searching mistakes. - Export a beta version of the application to make trials. - Evaluate the results
Identify and fix possible mistakes	- Test it with my device and others to see the final results and things to improve. - Improve those features.

Now let's specify each task, defined before in Table 1.

First of all, we will take care about those tasks related with the decision of which softwares will be used to develop the application like:
- Register, and get if it's needed, in different app development softwares such as Android Studio, Xamarin (iOS), Ionic dashboard, appery.io and *Mobincube*.
- Register into Heroku, Netlify, Google App Engine and Firebase.
- Try to make a new app with every one and see how it works.
- Compare the different options setting as a principal point the easiness and the range of options/features to develop the app.

With these three tasks I expect to be able to choose which one would be the best one to develop my project.
- From the comparison done before select the best platform to develop our app.

So, once decided, and although I have tested each one before to compare them, let's dedicate some time to know all the features of the software, for that I decided to:
- Search into the selected app building software website to deepen into its options.
- Search for more information and examples on articles or internet.
- Practise

Once I've got the app controlled, let's start with the design and build up of the structure of the application along all the phase 1 of the project:
- Decide how many screens it would have and the function of each one and create it..
- Design the content of each one: buttons, tabs...
- Choose the symbols and other aesthetic factors.

Parallel to this, it's important to know which bioinformatic tools I am going to include into my app, because it will determine how is distributed this app (screens, buttons, etc.). So I won't be able to develop the complete structure of the app without knowing all the content. Because of this, while I am developing the app, I will:

- Search into articles and asking researchers the most important (and diary useful) bioinformatic tools.
- Search the internet for this applications and check if they have the API code.
- Make a list with those selected.

Once I have this first bioinformatic tools list, let's discuss if they would be adequate, or if I forget some of them. For this, the plan includes to ask some researchers their opinion, considering, of course, the fact that the API code of each bioinformatic tool is required:

- Argue why the chosen tools are the the best to include in the application.
- Ask the researchers if something is wrong or they want something that is not considered yet.

With this second and definitive list of the bioinformatic tools needed, I will be able to finish the structure of the application, and it's planned to be before the end of the phase 1 of the project.

In the second phase of the project I will focus all my efforts in the functional features of my app. For this, the first step will be learn how works the API code, so I kept some time for this in my planning. To learn how it works, I will divide my learning in three parts, based in *Mobincube*, other example apps and some tutorials:

- Learn about API directly from *Mobincube*.
- View other apps examples working with API code.

Once I have learned about how API works, I will link the different bioinformatic tools with my app. This is without a doubt, the harder task on my project plan. The steps to get the connection look simple (for now), get the API code and include it in my app.

- Link the software, databases with the different screens obtained in the structural chores.

It's planned to get to this point before the end of the phase 2 of the project.

After link all the bioinformatic tools, I will assess my final product, for that, and in first place, I'll check all the features of the app in two ways: from the *Mobincube* platform itself and from the device use. This evaluation will be internal and external, as explained in the point 5.4 Project evaluation.

- Review the application searching mistakes.
- Export a beta version of the application to make trials.
- Evaluate the results

To finish, with the results of the evaluation I expect to be able to improve the final product, although this will be possible only if I have the time, so the last task is to correct the identified mistakes and improve those evaluated features.

- Test it with my device and others to see the final results and things to improve.
- Improve those features.

So, with this plan, I set intermediary stages of the project:

1. Selection of the app building software to work with.
2. Get the application builded (in structural terms).
3. Make the list of the principal bioinformatic tools with API that will be included in the application, considering the opinion of current researchers.
4. Establish the connections with the chosen utilities.
5. Get the application ready to use.

Now let's see how the different tasks are allocate along the project (Figure 2).
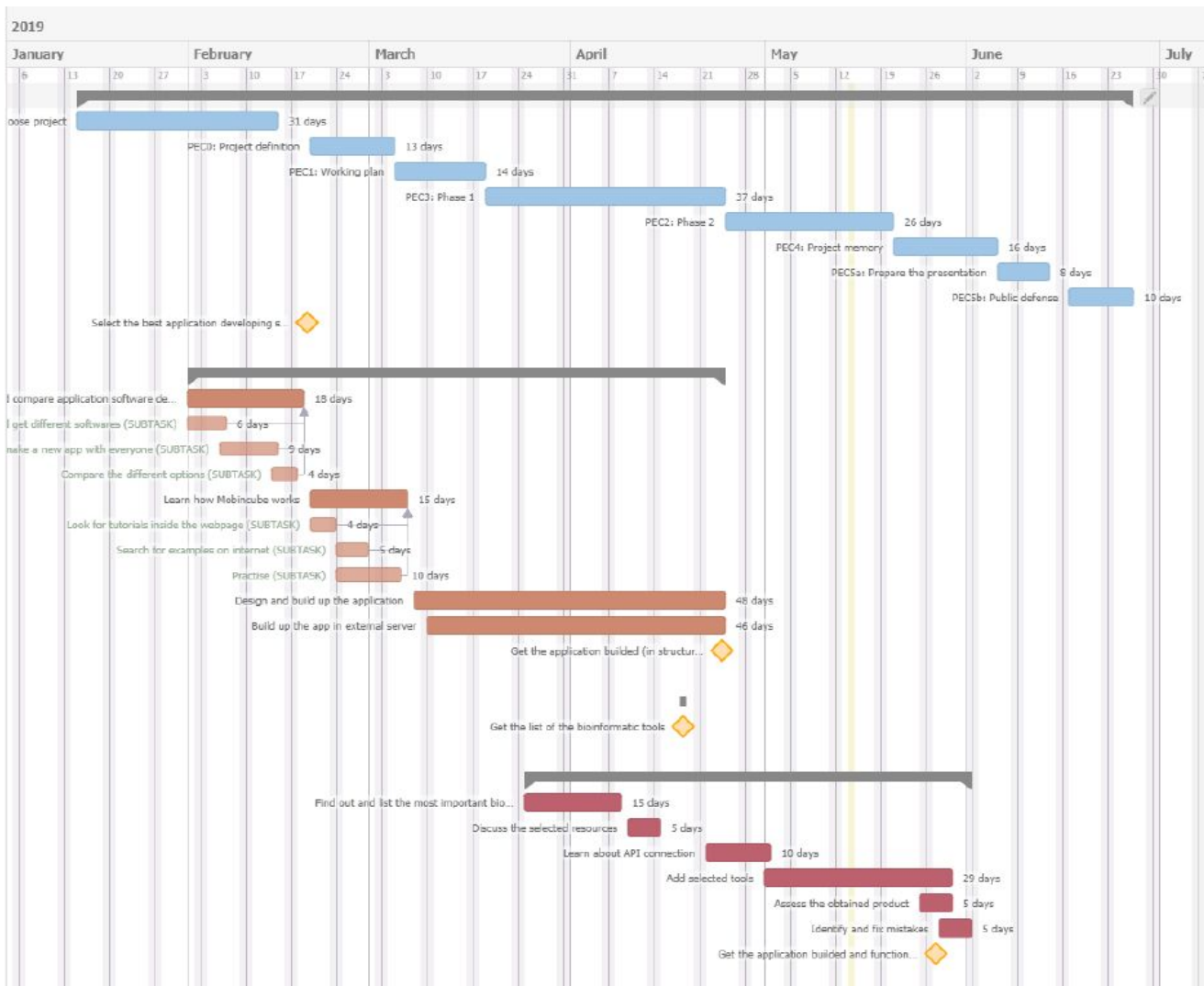
Figure 2: Calendar of the tasks, separated in paper tasks (in blue), representing the writing works like the project memory, structural tasks (in red), representing the build up of the app structure, and functional tasks (in garnet), representing the selection of bioinformatic tools and their connection with the app through API code. In yellow diamonds are represented the principal stages of the project. Image created by teamwork.

1.5. Obtained products

Now let's describe the results that we should expect from this project.

1. **Working plan:**

The first product obtained from this project is the working plan. In this planning are included a general description about the project, the objectives, methodology, tasks and timing.

2. **Memory:**

The second product obtained from this project is this written memory, where will be registered all the process from the beginning of the project.

3. **Product:**

But the real result of the project is the developed product itself, *Biopocket*.

The cross-platform application, after developed and tested, will be submitted in Android, iOS and WebApp format, available for all devices.

4. **Project evaluation:**

In the evaluation we find two parts, a self evaluation and an external evaluation.

The self evaluation is included in this project, especially over the developed application. In this assessment I will try to verify if all the objectives have been reached and rate the quality of the final obtained product.

The external evaluation is carried out with a *Google Questionar* sheet. This questionnaire was sent to the researchers who have advised me along the project and also have tried the application for a few days.

5. **Presentation:**

The last product generated in this master degree project is the presentation of the results in video record format.

In this video I will use an audiovisual presentation as a physical support and record my explanation over it to present all the obtained results.

1.6. Description of the memory chapters

This memory is divided in two chapters. Each chapter explain a key aspect on the app development and deployment or functionality.

❖ *Structural design of Biopocket*

This section deals with the structural design of Biopocket. This includes the main screen of the app, a reveiw of its different sections and the justification of each one, the differences between different platforms and the links between them.

❖ *Biopocket functional features*

This section deals with the functional features of Biopocket. This involves the process to choose the bioinformatic utilities connected into the app, the API code repositories of each tool, the function of the connected tools and the different sections, the differences between different platforms and the links between them.

❖ *Biopocket as a cross-platform application*

In this chapter figure the obtained product, all the necessary to get the biopocket application in each format and the used code.

❖ *Evaluation*

The evaluation has two sides, the self-evaluation and the external evaluation. In the first one I review the achievement of the project objectives and the quality of the obtained product. The second one is carried out through a Google Questionnaire sheet to verify is the application is useful.

❖ *Conclusions*

In this chapter I discuss the different sides of the project, its development and future possibilities.

2. Structural design of Biopocket

In this chapter is discussed the structural format and design of the application.

The structure of the app is important in order to be a real nexus for the different utilities integrated, because of this, the structure is very related to the functionality of Biopocket.

Is important to remember that the objective of the app is not only to join the bioinformatic tools in one site, Biopocket also seeks to be friendly for the users, so its structural design is the key to get this goal.

For this reasons, all the structure has been built to accomplish both objectives.

The mobile app, in iOS and Android format, has been developed with *Mobincube*, using its screens and blocks system and without coding, and the WebApp, with javascript as a programming language, and deployed in he *Netlify* server.

The app is organized in different sections accessible from the main screen of the app or through the menu bar located at the top of the screen. The different sections cover different steps of the research, as the search for scientific articles, browse different databases, perform some analysis, get citations to share, etc.

2.1. Main screen

The main screen of Biopocket (Figure 3) gives access to the different sections of the app.

Consists on a blocks menu with the name and a brief explanation of the app sections and a top bar. Through each block the user can access to the desired section. In the same way,the top bar, which is present in all the screens of the app, acts as a shortcut to move between the sections. It also has a link to the 'About Us' section, and allow the user to come back to the main screen named as 'Biopocket'.

Besides this, the main screen has the description of how Biopocket works and a list of interesting bioinformatic resources, some of them already included in the app and others tools that are not included.
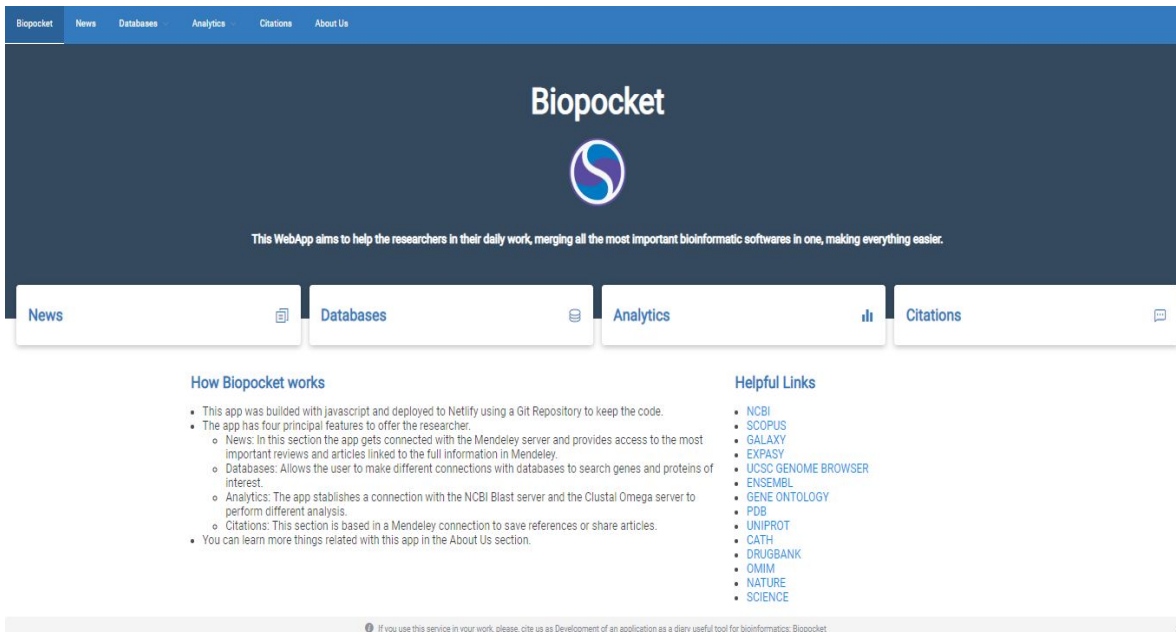
Figure 3: Main screen of Biopocket. On the top we have the top bar with the different sections, in the upper section the logo and principal objective of the WebApp followed by the four blocks of the different sections, the description of how Biopocket works and some helpful links.

2.2. Sections

As explained in the background, bioinformatics has different roles in bioscientific research, from those I highlighted 4:
- They can provide us a lot of information before and during the experiments.
- They are able to simulate some of them and predict their results.
- They are capable of comparing results through statistical analysis.
- They give us the way to share information and results.

The organization of the app in different sections is the way to include the different roles into Biopocket.

Of course, some of the roles could need specialized software to work, like for some simulations or statistical analysis, that can not be included in the app. That will be discussed in the 'Biopocket functional features' chapter.

The app has four sections, each one hold different bioinformatic tools in order to be related with the roles mentioned before. They are detailed in the following table (Table 2).

Table 2: Structural view of the different app sections and the function that will be located in each one.

NEWS	DATABASES
This section aims to provide information from articles before and during the experiments.	Platform that wants to provide information from different databases.
ANALYTICS	CITATIONS
To execute biological comparisons and perform analysis about them.	In this section the user can get citations of relevant articles, a way to share articles and information.

3. Biopocket functional features

In this chapter is discussed the functionality of the application.

The bioinformatic resources included are grouped in the different sections explained before. Let's review the election of each one, how perform the connection between the app and the server and also the function it performs.

3.1. Selected utilities

The choice of the utilities to integrate into Biopocket is based in two factors:

- Application Programming Interface (API):

In order to perform connections with a desired server, in a stable and easy way, the API code must be available in the documentation of the bioinformatic resource. The different API code repositories used to perform the linkage are detailed in the next section 'API code repositories'.

From the large list of bioinformatic tools that exist, the selection considering this factor are the following, grouped by section (Table 3).

Table 3: Selected utilities with available API code.

NEWS	DATABASES
- NCBI E-utilities - PMC Open Access - Elsevier/SCOPUS	- NCBI E-utilities (Entrez database) - PDB - UniProt - SwissModel - QuickGO
ANALYTICS	CITATIONS
- BLAST - Clustal - Galaxy server	- Mendeley

Must be considered the fact that, although a server has API code, it does not imply that all the functions are available through API connection.

From those selected, I interview specialized and non-specialized researchers to know which resources would they find more useful. By this way, I made sure not forget anyone.

- Surveys and interviews

The scientists interviewed indicated me which resources would be more interesting and useful for them. The surveys used are attached in the annex of this memory.

The profile of those interviewed belongs to researchers who use to work with bioinformatics and others that never work with them (Figure 4-A). Despite this, all of them know about the different tools and resources.

From the features offered by the different utilities that can be included in Biopocket, stands out the alignments, both gene and protein, which shows that the prefered section for the researchers is the analytics (Figure 4-B).

This makes clear that the analytics section should have alignment and biological analyzing tools. Of course, the inclusion of a pure statistical analysis software has no sense in this app because, if a researcher needs to carry out a deep statistical analysis, should use other resources like R software or similars.

Having a look to the other sections, the most rated feature in the article search is to have different journals (Figure 4-C). This was already considered, but looking deeper into the different API connections, I found that Mendeley offer also search in its database. The Mendeley database includes the most part of scientific journals and articles, so I decided to change the different databases selected (NCBI, PMC and Elsevier, also included in Mendeley) for a connection with Mendeley in order to have access to even more information.

Taking care of the results, in the database section the scientists find to have different databases as useful as the access to the information to the gene or protein of interest (Figure 4-D). So, this section has suffered no changes after the survey, but when looking to the API code, I found the PDB developer documentation so complicated, and considering that Biopocket would have a protein search software, thanks to the UniProt server connection, I decided to focus the efforts on the UniProt connection rather than in both.

I find the results of the question about the citations section surprising, I expected better rate about the citations manager or style. The most rated feature was allow the user to have a free search (Figure 4-E), so I will try it when I'll develop my application.

During the interview, I ask them about other ideas or approaches, and one of them told me that they use to work with the computer, so, it's a mess to be connected on the computer and be forced to pick up your smartphone to keep working or to find other resources. Because of this, I decided to design also a WebApp accessible from the computer in addition to the mobile app.

To finish, as reflected in the survey, the most part of them agree that Biopocket would be an useful tool and they will use it (Figure 4-F).

A

Do you usually work with bioinformatic resources?
8 respuestas

- ● Every day, it's my research field
- ● Every day, it's not my research field
- ● Sometimes (3-4 at week)
- ● Not usually (1 at week)
- ● Pratically never (few times at month)
- ● Never

B

Select from the following features which ones would you prefer to found in the app
8 respuestas

C

In the search of articles...

D

In the databases...

E

In the citations...

F

If you have this app in your device, imagine that all the discussed features before are included, would you use it?

8 respuestas
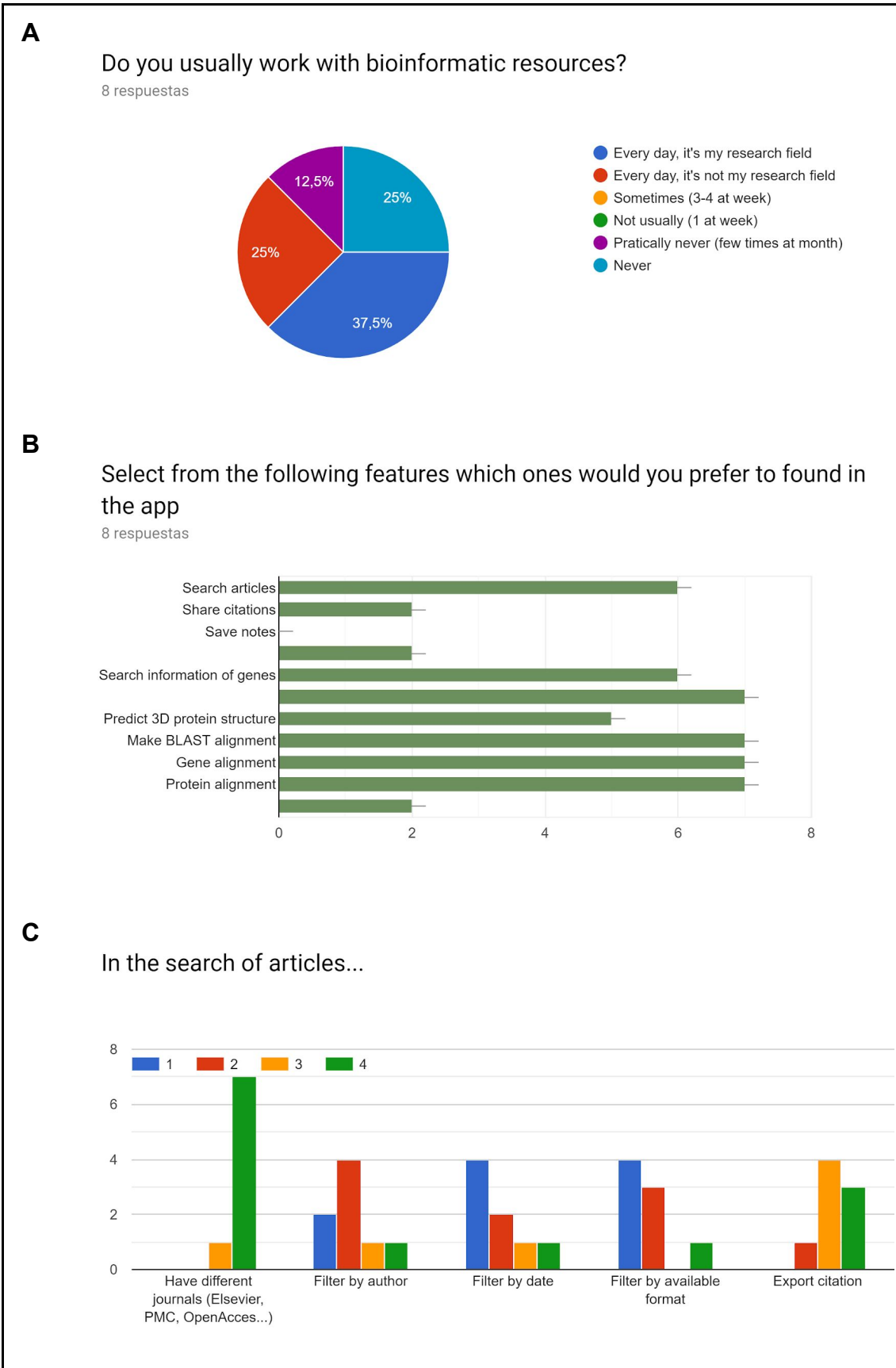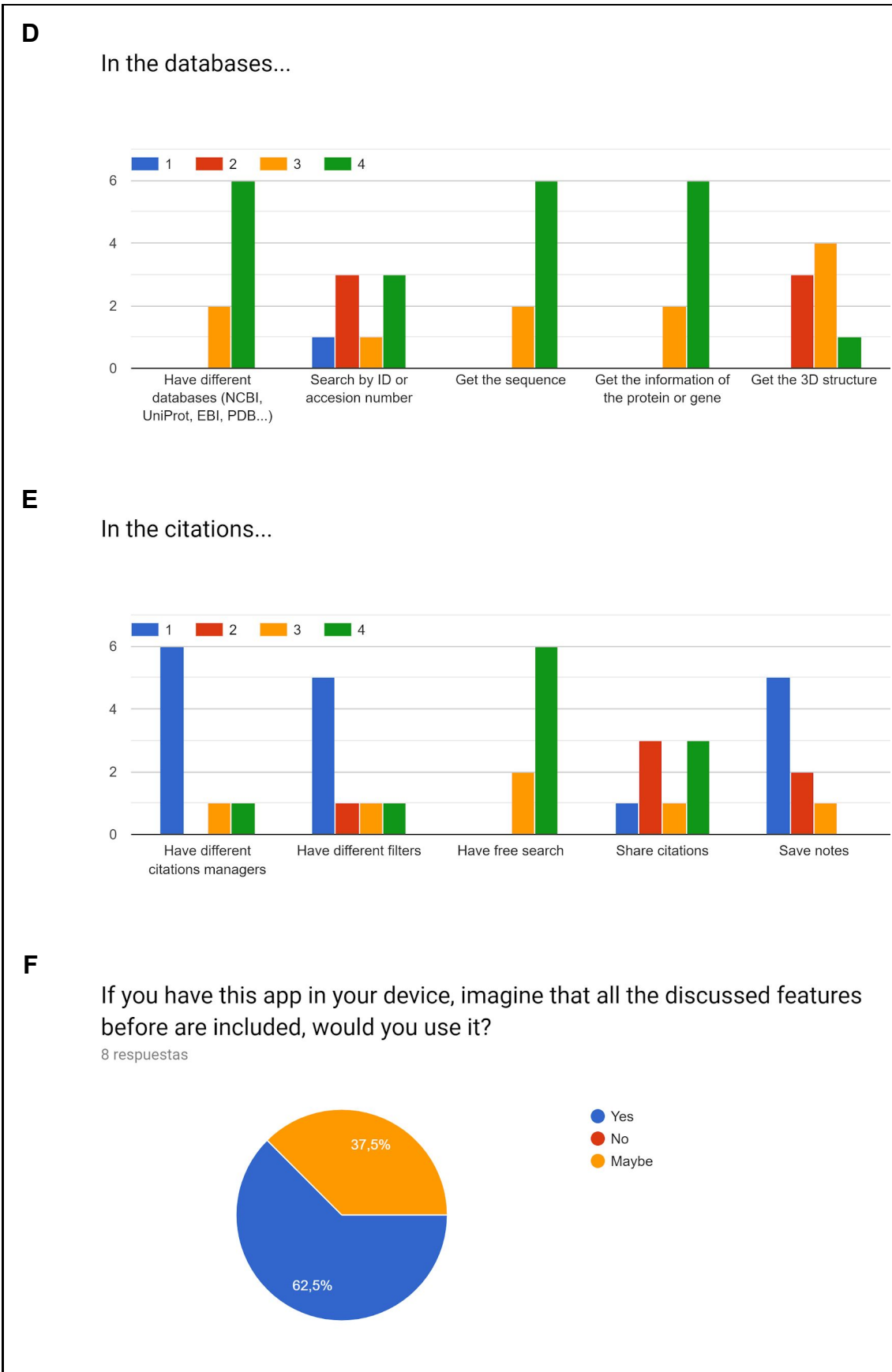
Figure 4: Results of the survey to researchers.

3.2. API code repositories

After the selection process of the different bioinformatic tools, those chosen must be connected with the app through their API code.

The list of the developer documentation of the different services is specified in the following table (Table 4). The links of each one is where they keep their API code and the instructions to establish a stable connection with the server.

Table 4: List of the developer documentation of the bioinformatic tools. For each one is specified the name, section to which it belongs and the link to its API code repository.

Mendeley	News and Citations
https://dev.mendeley.com/ https://api.mendeley.com/apidocs/docs https://dev.mendeley.com/methods/#introduction	
NCBI E-utilities	Databases
https://www.ncbi.nlm.nih.gov/books/NBK25501/	
UniProt	Databases
https://www.uniprot.org/help/programmatic_access	
BLAST	Analytics
https://ncbi.github.io/blast-cloud/dev/api.html	
Clustal	Analytics
https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/Clustal+Omega+Help+and+Documentation	

Some of the repositories detailed above on the Table 4 does not contain a pure API code. That's because some of them, like the NCBI E-utilities, doesn't work with a programmable interface, but with programmable links to send queries to the app.

All the details of the implemented code will be discussed in the 'Biopocket as a cross-platform application' chapter.

3.3. Functions

Each connection contributes to Biopocket in a different way, and it's clear that the combination makes a new and powerful bioinformatic tool.

Some things must be considered before analyze the functionality of the app and its connected services. It's important to remember that, as said before, the fact that a bioinformatic service has an available API code, does not mean that, through this API connection, the user would be able to use its full services. Moreover, we have to remind that all the softwares and hardwares will be used under free licenses and the connections to the servers have been established free of charge. This could limit the use of the *Mobincube* platform, *Netlify* server, and also some of the APIs connections.

Now let's describe the function of each resource integrated in Biopocket.

❖ Mendeley

The Mendeley connection has two functions: act as an article search tool for Biopocket and also as a citation generator.

The use of Mendeley in the citations section is clear because Mendeley is known to be one of the best citations manager, and, as discussed before, the implementation of Mendeley in the News section answer the need to get access to more scientific articles.

Due to its two functions, the Mendeley connection is the only one present in two different sections, in the news section and in the citation section. Of course, to use the Mendeley connection, the user must log in in Mendeley. This present one important advantage, because, all the searches carried out in Biopocket will be registered in the users Mendeley account. But also has one disadvantage, because the user should log in when entering each section, so, to simplify its use in Biopocket, the log in Mendeley page is located before entering to Biopocket (Figure 5), by this way, the API connection get established from the beginning and don't disturb the user while navigating through Biopocket.
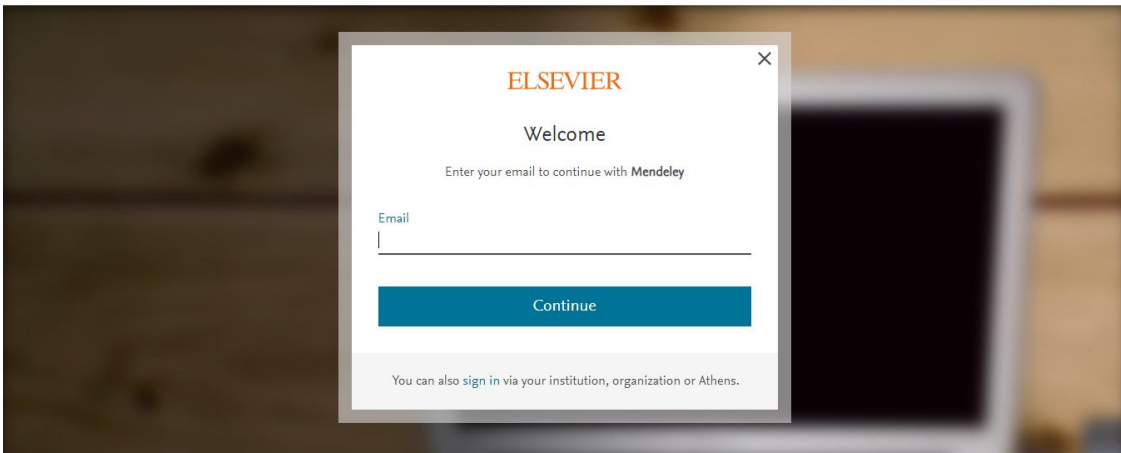
Figure 5: Log in screen when entering into Biopocket application. Its function is to establish connection with Mendeley server.

In the news section the Mendeley connection is used to search articles. The user can set different filters like search by author, title or source. The last one allows the user if is searching for an article, journal, book…

The output of the search consists in a different blocks. In each one the user can read the title and abstract of the article, the authors, year of publication and the type of source it is (Figure 6-A). If the user is interested in one article or want more information about it, just have to click on it, and a new tab will be opened in the web navigator with all the information related to the article on Mendeley (Figure 6-B). That includes, the publication date, journal, the citation of this article, the popularity, comments, related articles, sometimes the full article and more. The user has already logged in, so there's no need to introduce the credentials again and also can save the article directly to the users Mendeley account.

Figure 6-A: Screen output of the News section. In this example, the search is for 'hallmarks of cancer'.

Figure 6-B: Result of follow the link generated for each article in the News section. In this example, the first result for the search 'hallmarks of cancer' as indicated in Figure 6-A.

In the citations section the search has the same filters and, in this case, the result of the search provides the user a list of blocks with the title of the article, year and its reference in APA format. This reference is constructed with the different information given from a query to Mendeley database, that means that all of them are constructed combining the different information of the article: authors, journal, year, doi, etc. (Figure 7).

Some articles could not have all the information and because of this the reference could be incomplete.
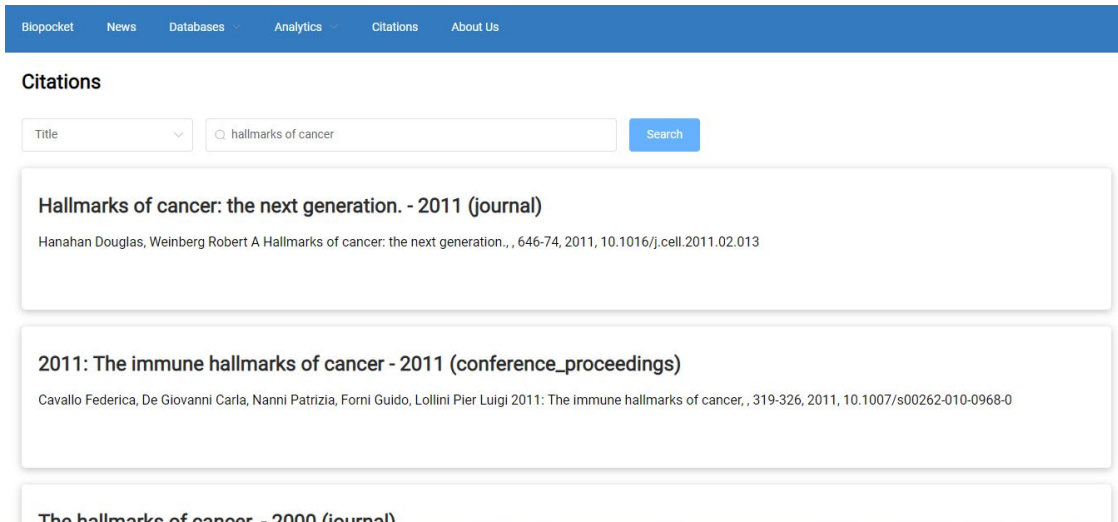
Figure 7: Screen with the output generated by the search on the Citations section.

The order of the results in the output, both news and citations, is based on the relevance of the articles.

❖ NCBI E-utilities

The NCBI E-utilities consist in eight different tools based in programmable links. Through this links you can send queries to the different databases of the NCBI, as PMC, protein or gene, among others. The nine E-utilities allows the programmer to search, fetch, post, link and get summaries.

In the case of Biopocket, the E-utilities used were the ESearch and the ESumary/Efetch, in order to get the information of genes from the gene and Entrez database.

The user only has to know the Gene ID, and with this information Biopocket will return the user the name of the gene, other possible alias, the description, location, annotation and the MIM number (for the OMIM database) (Figure 8)..

Figure 8: NCBI E-utilities output screen. Using the gene ID the search provide information about the gene.

❖ UniProt

The UniProt API is one of the most complete and easy programming in bioinformatics. The database provides a lot of information, because of this, I think is one of the most powerful resources in Biopocket.

It works with a search bar where the user only has to indicate the name of the protein to make a search. The output of this search includes the name of the protein, the classification method, its most relevant features and domains of the protein, its sequence and length (Figure 9-A).

Figure 9-A: Result of the search in the UniProt database. The most important of the output is the features of the protein and the sequence.

Moreover, each result has a cross-link to consult the gene information of the protein from the EBI database which includes CDS, FASTA, articles, and more (Figure 9-B).

Figure 9-B: European Nucleotide Archive page associated to the UniProt search. In this case, this result belongs to the first result for the 'akt' search done in the Figure 9-A.

❖ BLAST

The BLAST service in Biopocket acts as it does in the BLAST web service, allowing the user to carry out alignments of a known sequence comparing it with others in the desired sequence database. The user can search alignments of a gene or protein with gene or protein sequence database, which means that the user is able to compare gene to gene (blastn), translated gene to protein (blastx), protein to protein (blastp) and translated protein to gene (tblastn).

In the same way, Biopocket allows the user to make any of these comparisons by introducing a gene or protein sequence, by selecting the type of programm to be executed. The BLAST software in Biopocket is configured to perform alignments in the European Nucleotide Archive (ENA) database (for DNA samples) and in the UniProt KB database (for protein samples).

The output generated is a list where the user can get the identifier of the genes or proteins that match with the input sequence. This list contains the E-value of the match, and its score, calculated from its cover and the identity percentage of the match, which means the part of the sequence matched and the similarity between them (Figure 10-A).

After the list, the specific generated match is showed with the coverage and identity percentage (Figure 10-B).

Figure 10-A: Output list generated from the BLAST alignment tool.

Figure 10-B: Specific generated matches with the BLAST alignment tool.

❖ Clustal Omega

The Clustal Omega service in Biopocket allows the user to compare different sequences, both gene and protein.

The input are the sequences of interest, and the output generated is the resume of the analysis, with the name, number of sequences and time of execution, the percentage identity matrix (PIM) which represents the result of the comparison of the sequences and the phylogenetic tree which represents the evolutive distance between them (Figure 11).

Figure 11: Example of the Clustal Omega service. In this case, comparing two sequences of Akt protein from different organisms.

To finish this section, is important to explain why some resources why SwissModel, PDB or Galaxy are not included in Biopocket.

In the first case it was due to complications in the viewing of the 3D models obtained. The use of an specific external library as a viewer was difficult and I should have need more time to configure it. For the PDB database, as reasoned before, the API code was so complicated and so different from the others, so I was not able to perform stable connection with the server. For the last one, Galaxy, an amazon server was needed to deploy the connection (Taylor J. et al., 2011), and as explained before, Biopocket is developed free of charges, what means that, for now, this resource can not be included in the app.

4. Biopocket as a cross-platform application

Biopocket has been developed for different platforms, and this means that the structure,design or functionality could change between them.

This problem has been avoided thank to one of the features of *Mobincube* that allows to open a website inside the application. So, I programmed the iOS and Android app to open the deployed WebApp on *Netlify*.

So, the core of Biopocket is the WebApp located on *Netlify*, which contains all the code to establish the connections with the different bioinformatic tools. The *Mobincube* software acts only as a WebApp viewer.

The WebApp url is:
https://biopocket-app.netlify.com/

All the code deployed to *Netlify* is located in the following GitHub repository:
https://github.com/stejedorfelez/Biopocket-app

The demo for mobile (in iOS format) can be viewed here:
https://appetize.io/app/rt6vwrrm2u1jfpfd6erabz75r8?device=iphone6s&scale=75
&orientation=portrait&osVersion=11.4

The Android and iOS app can be downloaded from:
https://github.com/stejedorfelez/Biopocket-app/tree/master/Mobile%20applicatio
n

The mobile application is not available in the App Store yet, but yes in Google Play named Biopocket (Figure 12).
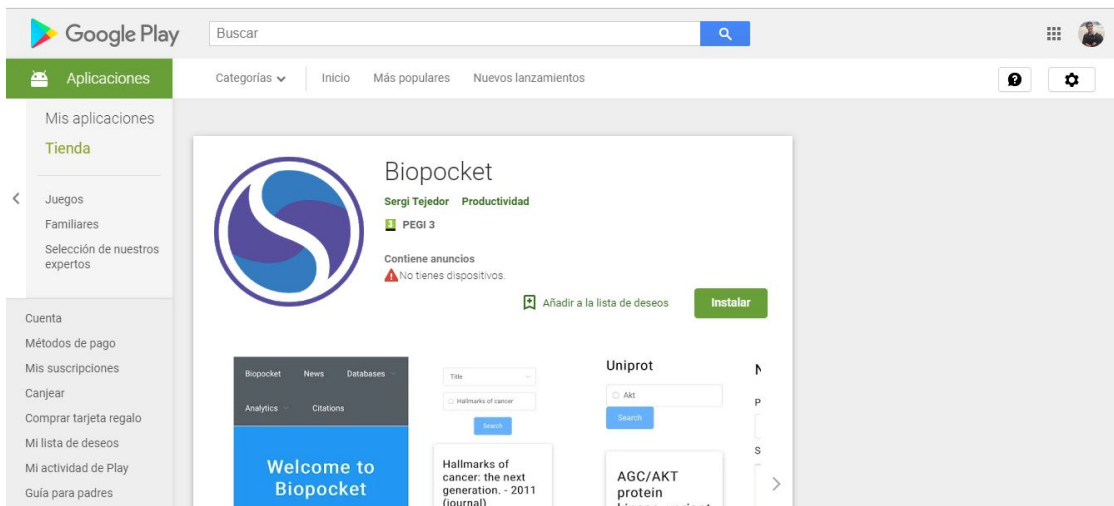

Figure 12: Biopocket in the Google Play market.

5. Evaluation

5.1. Self-evaluation

Evaluating the progress of the project, I must admit that, the fact that never worked before with mobile apps, javascript or API and neither had technical skills in this field, has delayed the development of Biopocket. This risk was considered from the beginning of the project.

Because of that, the phase 2 of the project wasn't as complete as I wanted, and not all the bioinformatic tools selected were integrated in Biopocket. Despite this, the development of the app was finished before the end of the project without problems. Probably, with more time, I could improve the interface of Biopocket and establish more connections with other resources to enrich the app.

Apart from this, focusing the analysis on the obtained product, I think that Biopocket is a really useful and innovative tool for researchers in this field, and opens the door for others to design similar applications making bioinformatics more accessible to everyone.

5.2. External evaluation

The external evaluation is carried out with *Google Questionnaire*. The surveys used are attached in the annex of this memory.

In the survey the researchers after use Biopocket, rate different aspects of the application, the results are represented in the Figure 13.
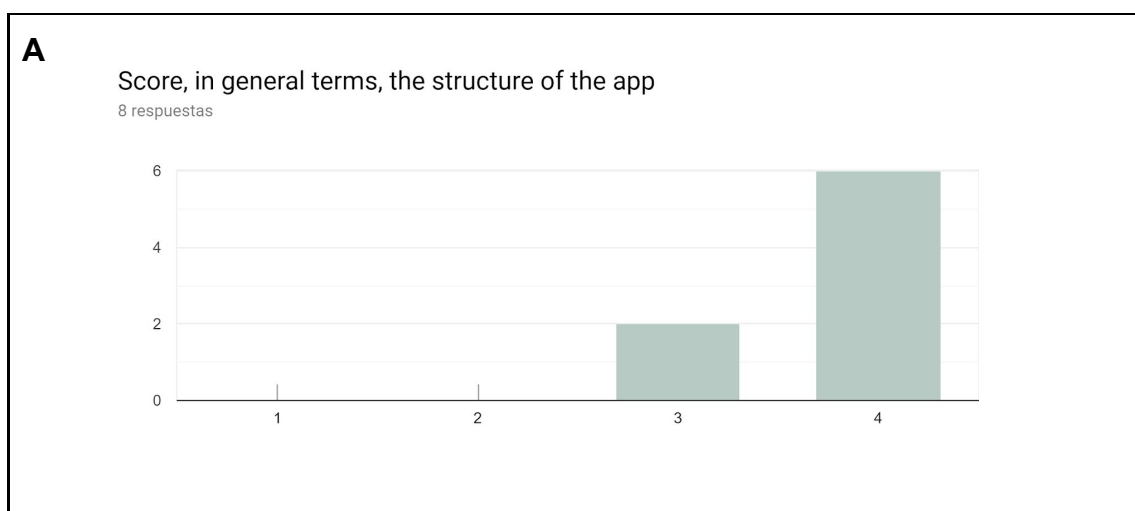
B

Score the search implemented system in the news section, considering the filters, matches and generated links to other sites

8 respuestas
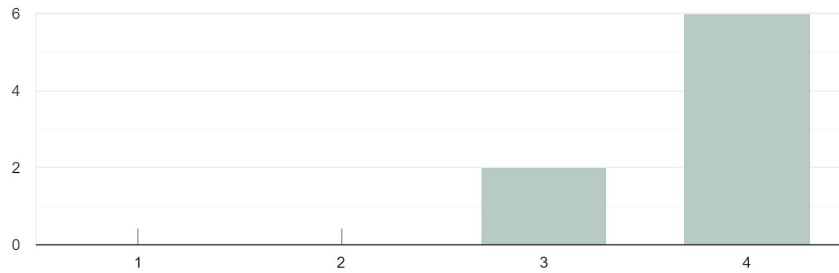
C

Score the search implemented system in the citations section, considering the filters and matches

8 respuestas

D

In your opinion, which section is the best resource of this app?

8 respuestas

- News
- Databases
- Analytics
- Citations

E

In your opinion, which section is the worst resource of this app?
8 respuestas

- News
- Databases
- Analytics
- Citations

87,5%
12,5%

F

Would you add any other bioinformatic tool to this app?
8 respuestas

- Yes
- No

25%
75%
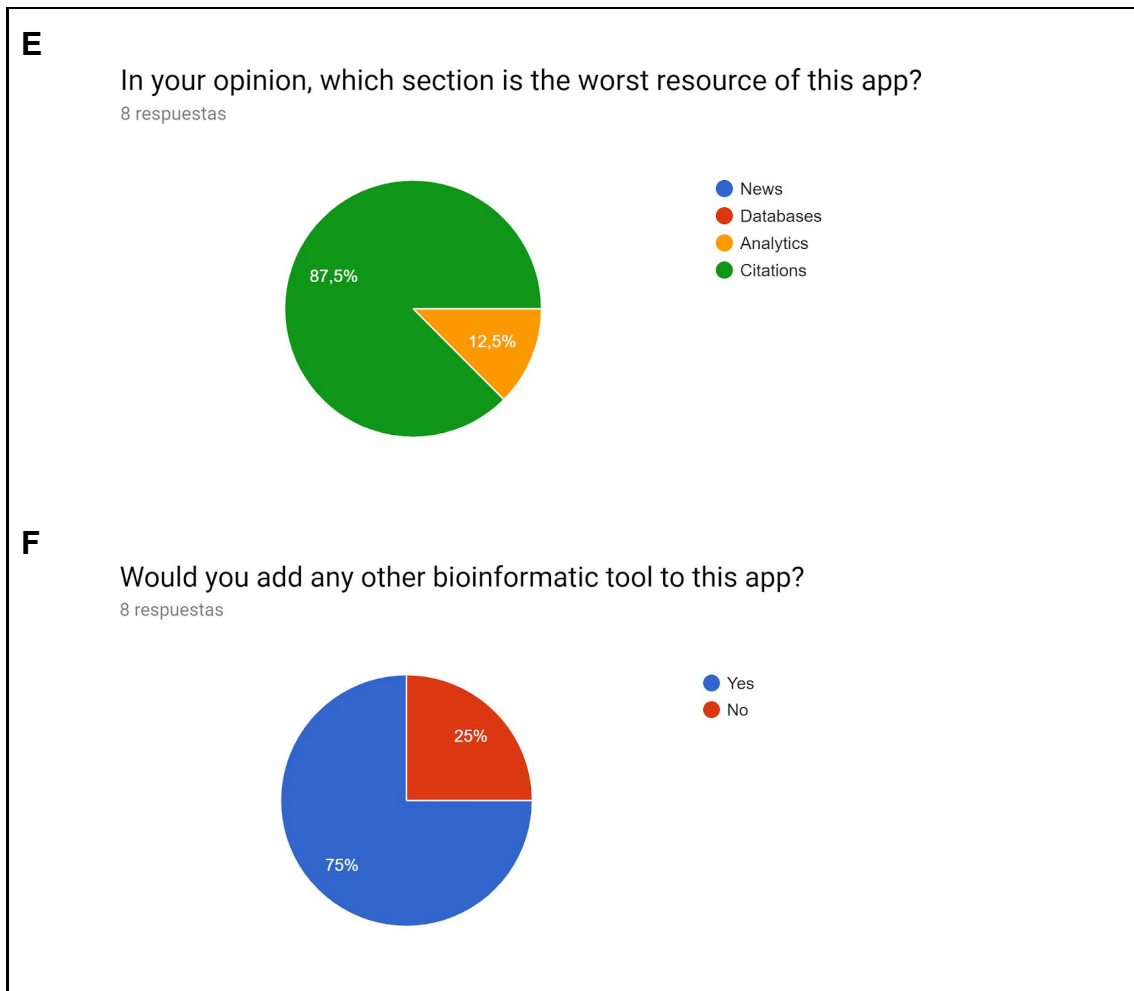
Figure 13: Results of the external evaluation survey

The external opinion rate the structure and design of the application very well (Figure 13-A). The most rated section of Biopocket is the news section (Figure 13-B and Figure 13-D) while the worst rated section is the citations section (Figure 13-C and Figure 13-F).

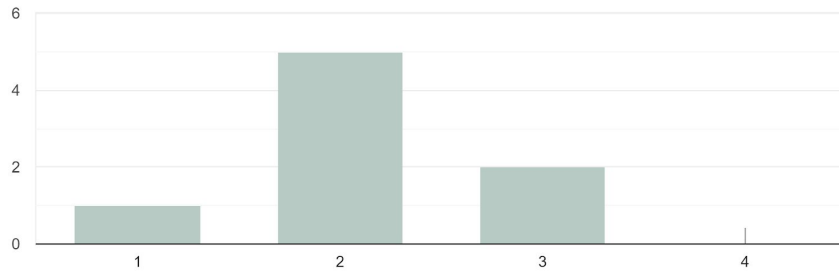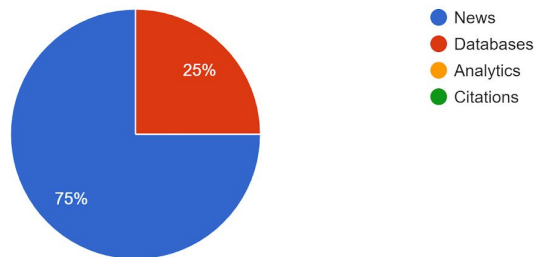The reason of this valoration is, for the news section, the link generated for each article that drives the user to Mendeley database with more information and, the most part of the times, the link to the full article.

For the citations section the users find that sometimes the generated citation is not fully correct, and the news section is linked to Mendeley, where they can get the citation, so the both sections could be merged in one.

The databases section is also well rated, some of the researchers point out that exists other databases, but, as discussed before not all of them can be included into the app.

For the analytics section some of them expect tools for statistical analysis, but, as said before, Biopocket is not a statistical software, for that is recommended to use other applications such as R software.

The commented about the abscense of some databases and the expected statistical analysis tools in Biopocket is reflexed in Figure 13-F, where the researchers demand more tools for the app.

After see the results of the survey I think that it's something positive this kind of apportations, because that means that Biopocket is useful for them and they want to improve its features and possibilities, so that make me realize that Biopocket could grow up and get even greater.

6. Conclusions

I can draw different conclusions from this project.

First of all, as I already said in this memory, I have never worked in the app developing field, so this project has give it to me the chance to do it. I think that the app development it is increasingly important, because of the evolution of mobile technology and the paper of devices in our day to day routine. How many apps do we use per day?

But this is not the only thing I learned during this project, this is only the coding part. If we focus in the field of the biological research, I have learned the relevance of bioinformatics, although that was clear to me, the fact that people who never use to work with this type of resources know them all, is clearly significant. Moreover, the search for bioinformatic utilities let me discover the wide range of tools that are developed, and this, confirm the real need of an application like Biopocket to integrate all them together.

The working plan has been accomplished with little variations in the calendar. This made that some objectives were not reached at the planified time, but later. The variations of the working plan were considered in the risks of this project and are related to my inexperience in the app development. This also made that the first planned methodology, based on *Mobincube,* was not enough to carry out the project. But after some research, the inclusion of *Netlify* as a server where locate the WebApp solved the problem.

In the same way, some resources have not been connected to Biopocket although having available API code, mainly due two reasons. First, the complexity of the code, and in second place, the lack of time.

Despite the problems, Biopocket has born as a cross-platform application, available for the most part of devices, and with some of the most used and powerful bioinformatic resources. Without a doubt, it is a big step for the bioinformatic tools, just because there is a unique platform for the information of the field.

Through the surveys carried out to researchers, specialized and non-specialized in the field, we can observe that Biopocket is really a helpful resource for our research, because made our day to day easier reducing the time spent in the development of a project.

This is why on a future I think that Biopocket should be improved by the introduction of more bioinformatic tools, maybe could start adding databases such as the protein data bank (PDB) or browsers like the UCSC Genome Browser.

With the time, the addition of a pure statistical section could be considered integrated with the help of the R shiny software.

Moreover, we must remember that Biopocket has been developed with free license hardware and software, which means that could be improved buying the license for some services if the application, in the future, is configured to earn some money.

7. Glossary

<u>In silico research:</u> In silico is a term related with all those processes carried out by computers. Focusing on biological research, in silico could refer simulations, analysis, modelling or others.

<u>API</u>: The application programming interface is the code with the specific language of a service. This means that is a set of definitions, protocols and tools for build code in order to establish a connection with this service. The programmer is provided with the building blocks and can develop the software for a web-based system, operating system, database or computer.

<u>WebApp:</u> A web application (WebApp) is a client-server program that can run in a web browser, allowing the interaction between the client and the server through its interface.

<u>Javascript:</u> A programming language that allows the user to execute actions in a dynamic website. This includes movement, interactions, connections with other services and more.

<u>Deploy:</u> To deploy an application is a process where the user copy, configure and enable an application to a base URL located on a server. Once the deploy is finished, the application becomes publicly accesible on the base URL.

8. Bibliography

¿Qué es Mobincube? (2019) [online] Barcelona: Mobincube. [Query: 27/02/2019]
<https://support.mobincube.com/hc/es/articles/204567849--Qu%C3%A9-es-Mobincube->

Afgan E., Baker D., Coraor N., Goto H., Paul I.M., Makova K.D., Nekrutenko A., Taylor J., "Harnessing cloud computing with Galaxy Cloud," Nature Biotechnology, Vol 29, Issue 11, 2011.

Bertrand Néron, Hervé Ménager, Corinne Maufrais, Nicolas Joly, Julien Maupetit, Sébastien Letort, Sébastien Carrere, Pierre Tuffery, Catherine Letondal, Mobyle: a new full web bioinformatics framework, *Bioinformatics*, Volume 25, Issue 22, 15 November 2009, Pages 3005–3011, https://doi.org/10.1093/bioinformatics/btp493

Common URL API (2019) [online] Barcelona: NCBI BLAST [Query: 20/03/2019] <https://ncbi.github.io/blast-cloud/dev/api.html>

Bioinformatics tools and software (2019) [online] Barcelona: Bioinformatics Software and Tools [Query: 09/03/2019]
<http://bioinformaticssoftwareandtools.co.in/bio_tools.php>

Entrez Programming Utilities Help (2010) [online] Barcelona: NCBI Entrez Programming Utilities [Query: 18/03/2019]
 <https://www.ncbi.nlm.nih.gov/books/NBK25501/>

Funciones avanzadas (2019) [online] Barcelona: Mobincube. [Query: 27/02/2019]
<https://www.mobincube.com/es/tour-generador-apps-moviles.html#Features>

Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., … Vincent, K. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, *11*(8). https://doi.org/10.1186/gb-2010-11-8-r86

Mariette, J., Escudié, F., Bardou, P., Nabihoudine, I., Noirot, C., Trotard, M. S., … Klopp, C. (2016). Jflow: A workflow management system for web applications. *Bioinformatics*, *32*(3), 456–458. https://doi.org/10.1093/bioinformatics/btv589

Mehmood MA, Sehar U, Ahmad N (2014) Use of Bioinformatics Tools in Different Spheres of Life Sciences. J Data Mining Genomics Proteomics 5:158. doi:10.4172/2153-0602.1000158

Mendeley developers portal (2019) [online] Barcelona: Mendeley [Query: 15/03/2019] <https://dev.mendeley.com/>

UniProt programmatic access (2019) [online] Barcelona: UniProt [Query: 15/03/2019] <https://www.uniprot.org/help/programmatic_access>

Sridhar, S., Higgs, B. W., & Althammer, S. (2018). Bioinformatics. In *Tissue Phenomics: Profiling Cancer Patients for Treatment Decisions* (pp. 131–156). Pan Stanford Publishing Pte. Ltd. https://doi.org/10.1201/9781351134279

9. Annex

Initial survey

Bioinformatics in your pocket

What would you think if you could handle bioinformatic tools with your device?

Sure you're thinking about searching articles, research about proteins or genes in different databases, study the tridimensional structure of a protein, submit a blast in the NCBI, making statystical analysis with Galaxy or sharing citations and notes.

Well, with the new mobile application Biopocket, all this is up to you in one touch.

*Obligatorio

About you
First of all, let's answer some questions about you and your relationship with bioinformatics.

1. **Do you work in bioscientific research?** *
   Marca solo un óvalo.

   ◯ Yes

   ◯ No

2. **Do you usually work with bioinformatic resources?** *
   Marca solo un óvalo.

   ◯ Every day, it's my research field

   ◯ Every day, it's not my research field

   ◯ Sometimes (3-4 at week)

   ◯ Not usually (1 at week)

   ◯ Pratically never (few times at month)

   ◯ Never

3. **Despite working more or less with bioinformatic tools, did you study and know the diferent available resources from bioinformatic field?** *
   Marca solo un óvalo.

   ◯ Yes

   ◯ No

4. **What do you understand for bioinformatic tools or resources?** *
   Selecciona todos los que correspondan.

   ☐ Those used to analyze the results of the experiments

   ☐ Those used to make predictions and simulations

   ☐ Those that allow to share articles, citations, comments and get in contact with other researchers

   ☐ Those that stock all the "bio" information (genes, proteins, diseases, etc.)

App sections

5. **Select from the following features which ones would you prefer to found in the app** *
Selecciona todos los que correspondan.

- ☐ Search articles
- ☐ Share citations
- ☐ Save notes
- ☐ Be able to connect with an external server
- ☐ Search information of genes
- ☐ Search information of specific proteins
- ☐ Predict 3D protein structure
- ☐ Make BLAST alignment
- ☐ Gene alignment
- ☐ Protein alignment
- ☐ Possibility to contact the app developer
- ☐ Otro: _____

6. **In case the app includes all the features listed in the last question, would you include something more to the app? Explain it** *

_____

_____

_____

_____

Rating the app
This app has four sections (Articles, Databases, Analytics and Citations), let's rate the options of each one:

7. **In the search of articles...** *
Marca solo un óvalo por fila.

	1	2	3	4
Have different journals (Elsevier, PMC, OpenAcces...)	◯	◯	◯	◯
Filter by author	◯	◯	◯	◯
Filter by date	◯	◯	◯	◯
Filter by available format	◯	◯	◯	◯
Export citation	◯	◯	◯	◯

8. **In the databases...** *
Marca solo un óvalo por fila.

	1	2	3	4
Have different databases (NCBI, UniProt, EBI, PDB...)	◯	◯	◯	◯
Search by ID or accesion number	◯	◯	◯	◯
Get the sequence	◯	◯	◯	◯
Get the information of the protein or gene	◯	◯	◯	◯
Get the 3D structure	◯	◯	◯	◯

9. **In the analytics... which procedure would you like to find into the app?** *

_____

_____

_____

_____

_____

10. **In the citations...** *
Marca solo un óvalo por fila.

	1	2	3	4
Have different citations managers	◯	◯	◯	◯
Have different filters	◯	◯	◯	◯
Have free search	◯	◯	◯	◯
Share citations	◯	◯	◯	◯
Save notes	◯	◯	◯	◯

Final questions:

11. **Something to add?**

_____

_____

_____

_____

_____

12. **If you have this app in your device, imagine that all the discussed features before are included, would you use it?** *
Marca solo un óvalo.

◯ Yes

◯ No

◯ Maybe

Evaluation survey

Biopocket evaluation

This questionnaire pretends to evaluate the functionality and structure of Biopocket, a new application that aims to facilitate the day to day research in bioinformatic research field.

*Obligatorio

Score and answer the next questions about the app

1. **Score, in general terms, the structure of the app** *
   Marca solo un óvalo.

1	2	3	4
◯	◯	◯	◯

2. **Do you think the structure is easy to work with and the sections are well implemented?** *
   Marca solo un óvalo.

1	2	3	4
◯	◯	◯	◯

3. **Score the search implemented system in the news section, considering the filters, matches and generated links to other sites** *
   Marca solo un óvalo.

1	2	3	4
◯	◯	◯	◯

4. **Score the database section considering the available options** *
   Marca solo un óvalo.

1	2	3	4
◯	◯	◯	◯

5. **Score the analytics section considering the available options** *
   Marca solo un óvalo.

1	2	3	4
◯	◯	◯	◯

6. **Score the search implemented system in the citations section, considering the filters and matches** *
Marca solo un óvalo.

	1	2	3	4
	◯	◯	◯	◯

7. **In your opinion, which section is the best resource of this app?** *
Marca solo un óvalo.

◯ News
◯ Databases
◯ Analytics
◯ Citations

8. **In your opinion, which section is the worst resource of this app?** *
Marca solo un óvalo.

◯ News
◯ Databases
◯ Analytics
◯ Citations

9. **Briefly, why?** *

_____

_____

_____

_____

_____

10. **Would you add any other bioinformatic tool to this app?** *
Marca solo un óvalo.

◯ Yes
◯ No

11. **If affirmative, which one?**

_____

_____

_____

_____

_____