12

# IndoorLoc Platform: A Web Tool to Support the Comparison of Indoor Positioning Systems

Raul Montoliu, Emilio Sansano-Sansano, Joaquín Torres-Sospedra, Óscar Belmonte-Fernández

*INSTITUTE OF NEW IMAGING TECHNOLOGIES, JAUME I UNIVERSITY, CASTELLÓN, SPAIN*

## 1 Introduction

Geolocation systems have been present during decades allowing navigation services that guide users by car, foot or bicycle, and many others such as evacuation services and social network services. The Global Navigation Satellite Systems (GNSS) are able to provide these services in outdoor environments, but in most of the situations people spend a significant portion of their time in indoor environments such as offices, undergrounds, shopping malls, airports, etc., where these satellite-based positioning systems do not work. This is one of the reasons why the development of new indoor positioning and navigation systems has attracted the attention of many researchers in the last years.

This research effort has achieved the development of many different indoor positioning technologies, being the ones based on Received Signal Strength Indicator (RSSI) fingerprinting (He and Chan, 2016; Wu et al., 2013; Han et al., 2014) among the most popular. This technique is based on the measurement of the intensity of the received radio signals of the emitting devices (beacons) that are available at a particular place, and on the comparison of this measurement with a previously built RSSI dataset (also known as radio map). In this scenario, a fingerprint is an RSSI feature vector composed of received signal values from different emitting devices or beacons, associated with a precise position. The similarity of the received signals (fingerprint) with some of the stored fingerprints can be used to guess the approximate position on the subject. This technique is becoming increasingly important for indoor localization, since Wi-Fi is generally available in indoor environments where GPS signals cannot penetrate, and the wireless access points (WAPs) can be used as opportunistic beacons. Other types of indoor localization beacons (Bluetooth, RFID, etc.) can also be used in conjunction with Wi-Fi access points or as a standalone positioning system.

**Geographical and Fingerprinting Data for Positioning and Navigation Systems. https://doi.org/10.1016/B978-0-12-813189-3.00012-5**

Many different approaches have been object of research and many papers have been published trying to solve this indoor localization problem. However, it is very difficult to compare results from different approaches, since every research presents its estimated results using its own experimental setup and measures, and it is very difficult to reproduce the particularities of every single experiment. In the *Pattern Recognition and Machine Learning* research fields, the common practice is to test the results of each proposal using several well-known datasets (García et al., 2009). This allows researchers to fairly compare different methodologies in the literature. For instance, the UCI Machine Learning Repository (Lichman, 2013) and the web *Kaggle* (Goldbloom et al., 2017) are two well-known examples in this sense. However, in the fingerprint-based indoor localization research field, there is a limited number of such kind of databases (Nahrstedt and Vu, 2012; Torres-Sospedra et al., 2014, 2015b; Talvitie et al., 2014; Barsocchi et al., 2016; Moayeri et al., 2016).

Many different approaches have been object of research and many papers have been published trying to solve this indoor localization problem. However, it is very difficult to compare results from different approaches, since every research presents its estimated results using its own experimental setup and measures, and it is very difficult to reproduce the particularities of every single experiment.

This chapter consists of an introduction to the *IndoorLoc Platform*,[1] a web tool to support the comparison and the evaluation of indoor positioning algorithms. The platform is a centralized website where researchers can do the following actions:

1. Access to a public repository of datasets for RSSI fingerprinting.
2. Upload indoor positioning estimations on experimental setups included in the platform.
3. Include the estimation results in a ranking.
4. Analyze positioning methods.
5. Interact with the platform in a user-friendly environment to test the algorithms and datasets included.

In order to show a real example of the platform usage, this chapter also presents a comparative study of the performance of two fingerprinting-based indoor localization methods included in the platform when using four of the datasets also included in the platform. All the experiments presented are easily reproducible using the tools included in the platform.

The two methods shown differ in the methodology used to solve the indoor localization problem. They are a deterministic-based and a probabilistic-based method. The four datasets differ in the type of scenario where data has been captured, as for instance: the number of samples, the size of the scenario, the density of the samples, etc.

[1] http://indoorlocplatform.uji.es.

A preliminary version of this chapter was published (as a conference paper) in Montoliu et al. (2017). This chapter provides additional details of the *IndoorLoc Platform*, to help the reader to be aware of the different possibilities of the proposed web platform.

The rest of the chapter is organized as follows. Section 2 reviews related work. Section 3 describes the main sections in which the platform is divided. Sections 4 and 5 explain the datasets and methods, respectively, included in the platform. Section 6 presents a set of experiments performed using the algorithms and datasets included in the proposed platform. Section 7 describes a real case of use of the usage of the platform during a fingerprinting-based indoor positioning course. Finally, the most important conclusions arisen from this work are presented in Section 8.

## 2 Related Work

As it has been said in Section 1, most of the indoor positioning methods found in the literature present the experiments using their own experimental setup. A second related problem is that those datasets are not made available to the research community, making it impossible to reproduce the presented results. Both issues make a fair comparison of localization methods developed by different groups not feasible in a rigorous manner, since scenarios may change in an uncontrolled way.

A better way to compare positioning algorithms is to use the same experimental setup, and for that purpose, the use of a repository of prerecorded data in a large variety of buildings and contexts can be very useful. Some good examples of data repositories in the machine learning community are the UCI Machine Learning Repository (Lichman, 2013) and *Kaggle* (Goldbloom et al., 2017), both created for evaluating machine learning algorithms with common databases.

Another alternative are competitions where several research groups should prepare their methods to obtain the best results using a common experimental setup, or even the same prerecorded data. Some examples of competition are: Microsoft-IPSN (Lymberopoulos et al., 2014, 2015, 2016, 2017), EvAAL (Potortì et al., 2015), and EVARILOS (Lemic et al., 2015). The first off-site indoor location competition was the third track of the EvAAL-ETRI Indoor Location competition (Torres-Sospedra et al., 2017b), called *Wi-Fi fingerprinting in large environments*, which was held during the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN'15). In this event, the competitors had access to the *UJIIndoorLoc* (Torres-Sospedra et al., 2014) dataset, that has been included in the proposed platform. A similar competition was held in the Seventh International Conference on Indoor Positioning and Indoor Navigation (IPIN'16) (Torres-Sospedra et al., 2017a), where the dataset used was more challenging, since data provided by all sensors embedded in typical smartphones was included, acquired by different people moving in different types of buildings.

One of the main problems of such competitions is that when they finish, researchers cannot continue improving their methods. In addition, the different datasets are located in different web pages. The proposed web platform is focused on providing a common place for researchers to access to fingerprint-related datasets. Another of the main objectives is to provide a continuous competition without deadlines. Therefore, researchers will not have time restrictions to test their methods and submit their results to the platform.

The most similar work to the *IndoorLoc Platform* is Lemic et al. (2015), where the authors presented a web platform for evaluation of RF-based indoor localization algorithms with two core services: one focused on the storage of raw data and the other focused on automated calculation of metrics for performance assessment. They also include an SDK for convenient access to the platform from MATLAB and python. The two first characteristics are included in the proposed web platform. The SDK is not needed in our case, since users can directly interact with the web platform to upload their results. The main differences of the proposed web platform with respect to Lemic et al. (2015) are as follows:

1. It is more focused on fingerprinting methods.
2. It also includes a dashboard section where researchers can make experiments using the methods and datasets included in the platform in a user-friendly environment.
3. There is a ranking section where researchers can check the accuracy of their method against the methods of other researchers. In addition, the proposed web platform has been designed in order to easily upload new methods and datasets.

The *IndoorLoc Platform* has been designed with a state-of-art visual style and with a user-centered interface making the access to all the elements of the platform very intuitive. For instance, the home web page (see Fig. 1) directly presents the main sections of the platform. Another example is that users can download a dataset or upload a result with just a few mouse clicks. The platform is also responsive, and will automatically adapt to the device screen used to access it.

In addition, the platform has a high formative component, because even a user without programming knowledge can interact with the algorithms and datasets included. Although, it will be the users with a high programming skill who will be able to get a better advantage of the platform because, probably, they will be able to improve the results that can be obtained with the algorithms included in the platform.

## 3 Overview of the Platform

Fig. 1 shows the homepage of the platform. The homepage displays a summary of the contents of the four main sections in which the platform is structured, while the menu in the upper section of the platform allows access to each one of these sections.

The four main sections of the platform are as follows:

- **Datasets**: This section is a repository of several datasets stored in the platform. These datasets are available to download so users can use them in their own experiments.

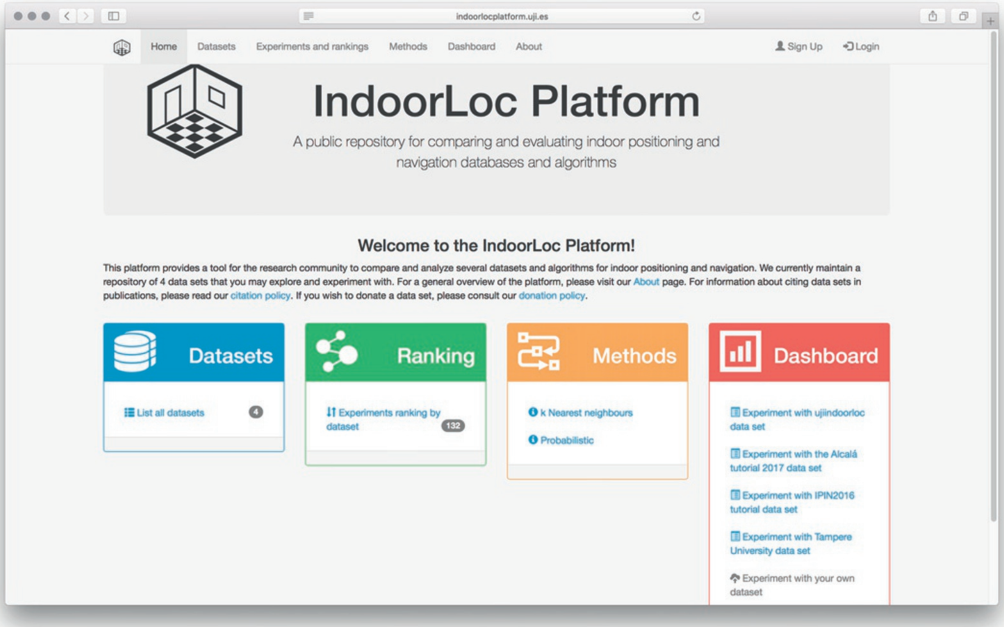


**FIG. 1** Homepage of the IndoorLoc platform.

- **Ranking**: In this section, users can upload the results of their own algorithms to obtain an estimation of the accuracy of their methods when using the datasets included in the platform. In addition, the results can be included in the ranking, where the best results of each dataset are showed sorted by accuracy.
- **Methods**: This section presents a set of well-known algorithms so users can study their implementation.
- **Dashboard**: In this section, users can test the algorithms included in the platform, using some of the datasets included, in a user-friendly environment.

These sections are briefly described in the next subsections.

## 3.1 Datasets

Fig. 2 shows the datasets section of the platform. The *Datasets* section displays the basic information about all the datasets included in the platform. In addition, the links to download all the files related to each dataset are also included.

Each dataset can be composed of up to four files:

- **Dataset info**: A *pdf* file with information and features about the dataset. The description includes the name of the donors, the contact information, general

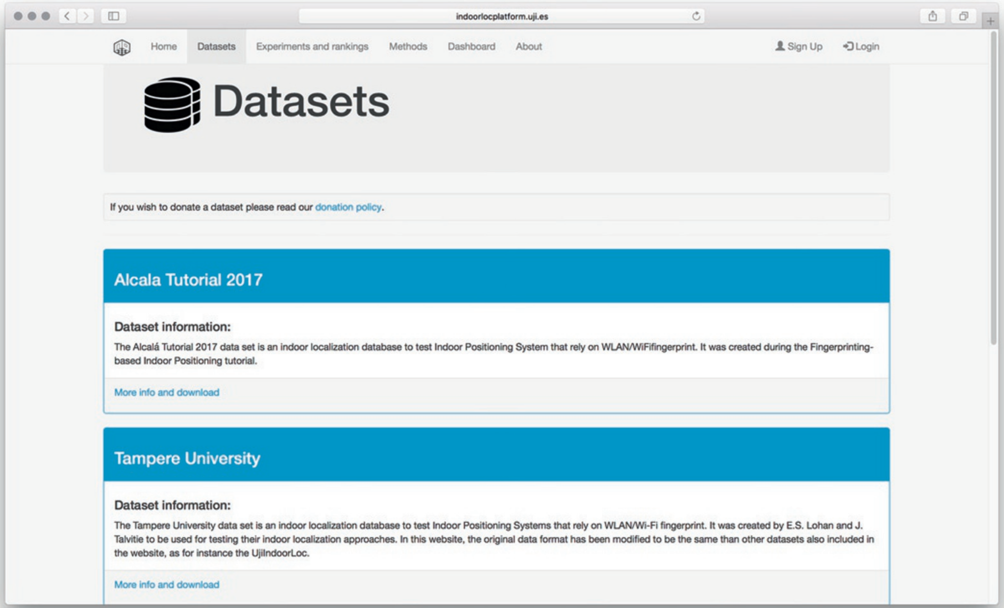


**FIG. 2** Datasets section of the IndoorLoc platform.

information about the dataset, a description of the files included, the attributes description, the format of the result file, and the citation request.

- **Training set**: A file with the samples to be used to train the localization models. It includes the localization of the samples.
- **Validation set**: This file is similar to the training set file, and also includes the localization of the samples. Should be used to assess the performance of the localization model created using the training set data.
- **Test set**: This file is also used to assess the performance of the localization model, but does not include the actual localization of the samples, only its fingerprint. To obtain an estimation of the accuracy of the model, users can run their methods to obtain an estimation of the localization of the samples of the test set, and then upload their results to the platform to get an evaluation of the performance of the model. The true localization of the samples is stored in the platform.

The training, validation, and test files have a *comma-separated values* (CSV) file format. The three first files (info, training, and validation) are accessible to everyone. Only registered users are allowed to access the test file. Not all the datasets included in the platform have a validation set. In that case, users can use techniques as cross-validation (Bishop, 2006) to assess the performance of the localization model generated.

At the moment of writing this chapter, there are six different databases included in the platform. Four of them are related to the Wi-Fi fingerprinting indoor localization problem. They are briefly described in Section 4.

Registered users can upload their own datasets following the instructions provided by the platform. Before to be definitively added to the platform, each new dataset is rigorously examined by the administrators of the platform to ensure that it has the required quality.

## 3.2 Ranking

One of the main objectives of the *IndoorLoc* web platform is to provide a tool to the indoor localization community to compare their methods using well-known datasets. This section has been devoted to this purpose. For each dataset, a list of the best methods, according to a figure of merit, is shown.

Registered users are allowed to upload the results of their methods following the instructions included in the description of the dataset (see Section 3.1). Once the results file has been uploaded, the platform calculates the figure of merit for this dataset using the estimated locations provided by the user and the ground truth internally (and privately) stored in the platform. After the figure of merit is calculated and displayed, users have the choice of including or not the result in the ranking.

Each entry in the ranking has a description field, provided by the user, showing info about the experiment performed to obtain such result, e.g., the parameters used or the algorithm details.

Fig. 3 shows the ranking page for the *IPIN2016 Tutorial* dataset. At the moment of writing this text, the ranking is composed by two experiments performed by the same user. According to the notes written by the user, the result of the leader was obtained using the probabilistic method and the one in the second position using a knn algorithm.

Fig. 4 shows the ranking for the *UJIIndoorLoc* dataset. In this case, the four best results obtained in the third track of the EvAAL-ETRI Indoor Location competition (Torres-Sospedra et al., 2017b), where this dataset was used, have been manually introduced by the web creators to give a baseline reference.

## 3.3 Methods

This section shows some basic information about the methods included in the platform. This information consists on the explanation of the method though R[2] source code using comprehensible examples. In addition, links to the Dashboard section, where users can test these methods, are also included Fig. 5 shows the method webpage.

At the moment of writing this text, two methods have been included in the platform: deterministic-based (Bahl and Padmanabhan, 2000) and probabilistic-based (Youssef and Agrawala, 2005). They are described in Section 5.

[2]https://www.r-project.org/.

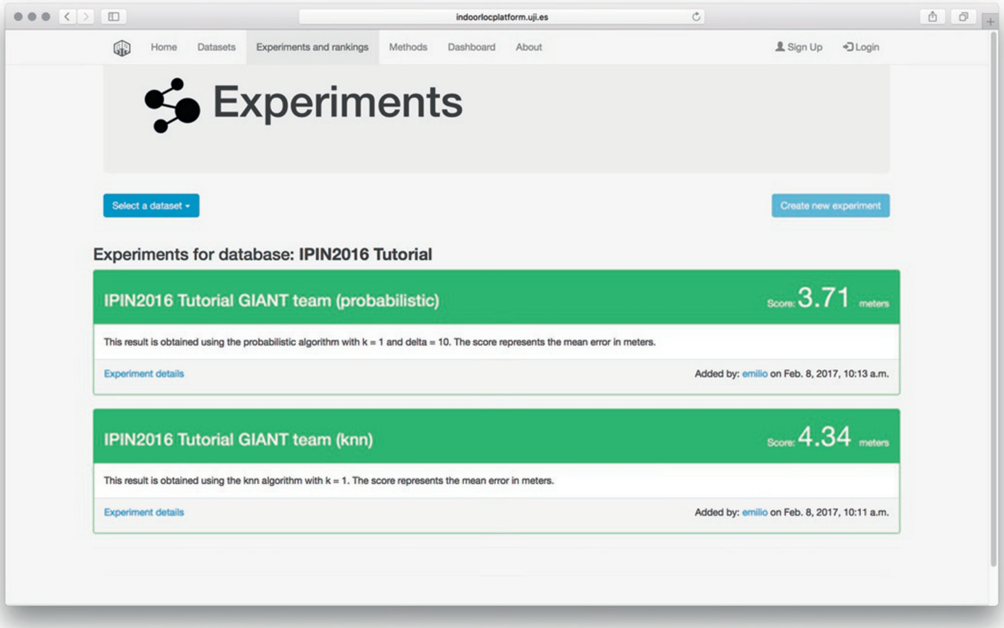


**FIG. 3** Ranking webpage of the *IPIN2016 Tutorial* dataset. Two experiments have been included in the ranking, the first one (according to the notes written by the contributor) using a probabilistic-based algorithm and the second one using a knn-based method.

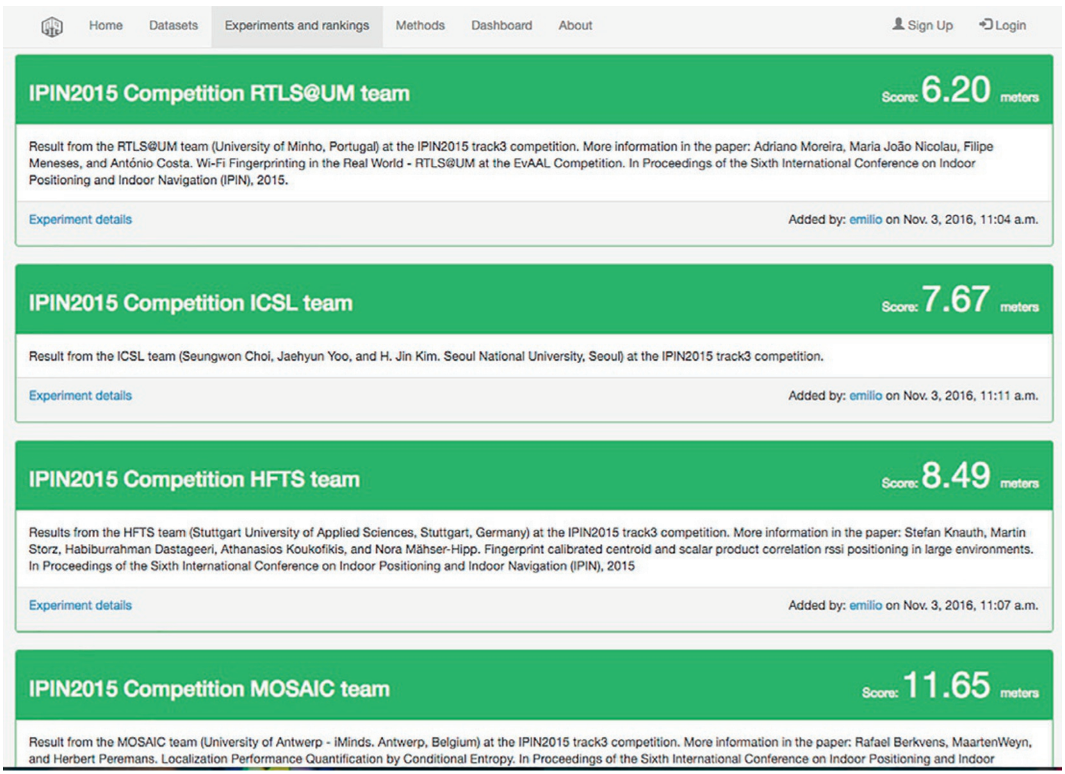


**FIG. 4** Ranking webpage of the *UJIIndoorLoc* dataset.

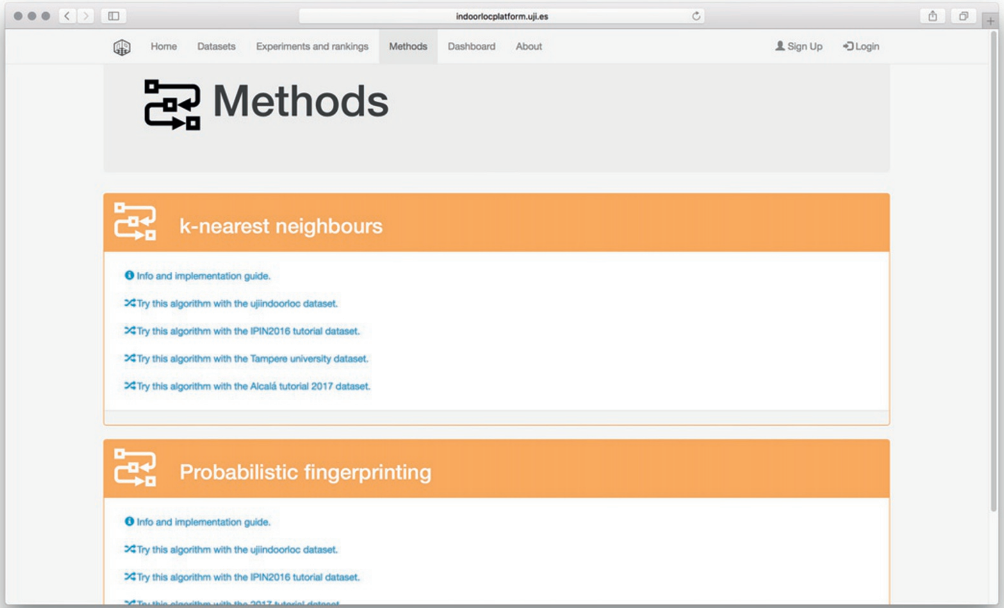


**FIG. 5** Methods section of the IndoorLoc platform.

To add new methods to the platform, users must contact with platform administrators. Similarly to the dataset case, each new method is rigorously examined by the administrators of the platform to ensure that it has the required quality.

## 3.4 Dashboard

In the Dashboard section, users can test the methods included in the platform, using the datasets also included in the platform, in a user-friendly interface. Fig. 6 shows an example of a dashboard for the *UJIIndoorLoc* dataset. In particular, the user selected the building 0, the floor 1, the validation set (to estimate the locations), and the deterministic method. After clicking on the *Estimate error* button, the platform internally estimates the location of the validation samples and calculates some statistics, as the mean and the median of the estimation error. It also shows one figure with the error histogram and density, and another figure with the empirical cumulative density function.

Registered users are allowed to use their own dataset using the methods included in the platform. This dataset must be formatted using a set of rules specified in the web platform.

The platform can be easily improved adding more measurements to estimate the methods performance and with other kind of figures, thanks to the R Shiny environment.

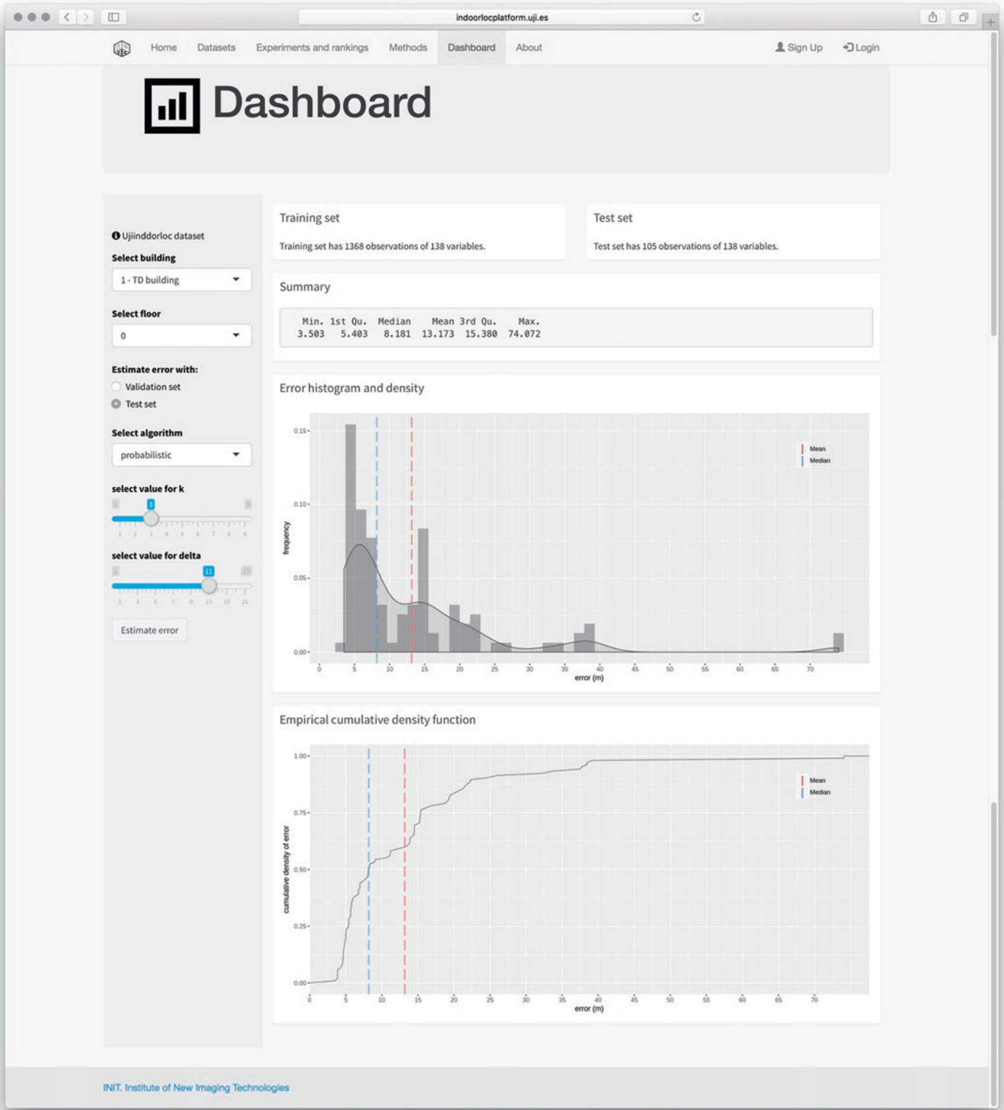


**FIG. 6** Example of the *Dashboard* section of the platform.

### 3.5 Implementation Details

The platform has been implemented using these open source tools:

- **Django**[3]: A Python web framework to build the web application. Django follows the model-view-template (MVT) architectural pattern, and allows rapid development of database-driven websites.
- **Shiny**[4]: An R package that eases the building of interactive web *apps* using R code. The Shiny server hosts the *apps* on the platform that run embedded in the dashboard page.
- **RMarkdown**[5]: Documents created with the R Markdown technology are fully reproducible, and use a notebook interface to weave together text and code to produce elegantly formatted output. R Markdown uses multiple languages including R, Python, and SQL.
- **ipft R package**[6]: This R package includes algorithms and utility functions for indoor positioning using fingerprinting techniques. These functions are designed for manipulation of RSSI datasets, estimation of positions, comparison of the performance of different models, and graphical visualization of data.
- **Apache**[7]: Open source HTTP web server that works on Unix-like systems (BSD, GNU/Linux, etc.), Microsoft Windows, Macintosh, and other platforms, and provides HTTP services in sync with the current HTTP standards.

## 4 Datasets Included in the Platform

At the moment of writing this chapter, the platform includes six different datasets, four of them dedicated to the Wi-Fi fingerprinting problem. They are briefly introduced in the next sections.

### 4.1 Wi-Fi-Based Datasets

In the four Wi-Fi-based datasets, Wi-Fi fingerprints are characterized by the detected WAPs and the corresponding RSSI. The intensity values are represented as negative integer values near to $-100$ dBm (extremely poor signal) to 0 dBm. The positive value 100 is used to denote when a WAP was not detected. Tables 1–4 show a summary of the main characteristics of each dataset.

[3]https://www.djangoproject.com/.
[4]https://shiny.rstudio.com/.
[5]https://rmarkdown.rstudio.com/.
[6]https://cran.r-project.org/web/packages/ipft/.
[7]https://httpd.apache.org/.

**Table 1** Main Characteristic of the *UJIIndoorLoc* Dataset

| | |
|---|---|
| Number of buildings | 3 |
| Number of floors | 4–5 |
| Number of WAPs | 520 |
| Number of training samples | 19,937 |
| Number of validation samples | 1111 |
| Number of test samples | 4900 |

**Table 2** Main Characteristic of the *IPIN2016 Tutorial* Dataset

| | |
|---|---|
| Number of buildings | 1 |
| Number of floors | 1 |
| Number of WAPs | 168 |
| Number of training samples | 927 |
| Number of validation samples | 0 |
| Number of test samples | 702 |

**Table 3** Main Characteristic of the *Tampere* Dataset

| | |
|---|---|
| Number of buildings | 2 |
| Number of floors | 3–4 |
| Number of WAPs | 390–354 |
| Number of training samples | 1478–583 |
| Number of validation samples | 0 |
| Number of test samples | 589–175 |

**Table 4** Main Characteristic of the *Alcalá2017 Tutorial* Dataset

| | |
|---|---|
| Number of buildings | 1 |
| Number of floors | 1 |
| Number of WAPs | 152 |
| Number of training samples | 670 |
| Number of validation samples | 0 |
| Number of test samples | 405 |

#### *4.1.1 UJIIndoorLoc*

The UJIIndoorLoc (Torres-Sospedra et al., 2014) database covers three buildings of Universitat Jaume I[8] (Spain), with four or more floors and an area of almost 110,000 $m^2$. It can be used for classification, e.g., actual building and floor identification, or regression, e.g., actual longitude and latitude estimation. It was created in 2013 and 2014 by means of more than 20 different users and 25 Android devices. The database consists of 19,937 training/reference records and 1111 validation records. There is also a test file where the ground truth is not accessible.

The 529 attributes contain the Wi-Fi fingerprint, the coordinates (latitude, longitude, and floor) and Building ID, and other useful information such as the particular space (offices, labs, etc.) and the relative position (inside/outside the space) where the capture was taken, information about who (user), how (android device and version) and when (timestamp) Wi-Fi capture was taken, among other information. During the database creation, 520 different WAPs were detected. Thus, the Wi-Fi fingerprint is composed of 520 intensity values.

This dataset was used in the off-site track of the EvAAL-ETRI Indoor Localization Competition which was part of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN'15) (Torres-Sospedra et al., 2017b). The best results obtained in the competition have been included in the platform in the corresponding ranking.

Since the particular implementation of the localization methods included in the platform assumes that all the samples are in the same building and floor, the complete dataset has been divided into 11 different datasets.

#### *4.1.2 IPIN2016 Tutorial*

As an alternative of the *UJIIndoorLoc* dataset, the *IPIN2016 Tutorial* dataset is focused on the study of a small scenario. In particular, it covers a corridor of the School of Engineering of the University of Alcalá[9] (Spain). It is the place where a tutorial on Wi-Fi fingerprinting was held during the IPIN2016 conference. The database consists of 927 training/reference records and 702 test ones. The 177 attributes contain the Wi-Fi fingerprint (168 WAPs), the coordinates where it was taken, and other useful information.

#### *4.1.3 Tampere University*

This database (Cramariuc and Lohan, 2016) covers two building of the Tampere University of technology[10] (Finland), with four and three floors, respectively. In the first building, there are 1478 training/reference records and 489 test ones. The 312 attributes contain the Wi-Fi fingerprint (309 WAPs) and the coordinates (longitude, latitude, and height). In the second building, there are 583 training/reference records and 175 test ones. The 357 attributes contain the Wi-Fi fingerprint (354 WAPs) and the coordinates (longitude,

[8]http://www.uji.es.
[9]https://www.uah.es/es/.
[10]http://www.tut.fi/.

latitude, and height). An important difference of this dataset, with respect the *UJIIndoorLoc*, is that in the former there is just one sample in each training location, while in the latter the number of samples is between 10 and 30.

Data from the two buildings can be considered as two separate datasets, with no relationship between respective WAP labels and real access points MAC addresses, meaning that two columns with the same WAP name in either dataset may be assigned to different access points.

Similarly to the *UJIIndoorLoc* this dataset has been divided into seven different datasets.

#### *4.1.4 ALCALA2017 Tutorial*

This dataset was created during the 2017 Fingerprinting-based Indoor Positioning tutorial held in the School of Engineering of the University of Alcalá. Data was acquired in the same corridor than the *IPIN2016 Tutorial* dataset. The main differences between both datasets are: (1) a thinner grid was used to capture training data; (2) some users made mistakes labeling the training fingerprints. These errors have not been eliminated since it is a situation that can occur in a real scenario. Users should take into account this situation in their methods.

The database consists of 670 training/reference records and 405 test ones. The 154 attributes contain the Wi-Fi fingerprint (152 WAPs) and the coordinates where it was taken.

### 4.2 AmbiLoc Dataset

The *AmbiLoc* dataset (Popleteev, 2017) is a collection of ambient radio fingerprints, collected in multiple predefined locations across several testbeds. Instead of Wi-Fi signals, the *AmbiLoc* datasets deals with ambient signals of opportunity, such as those from broadcasting TV and FM radio stations or GSM networks, that are almost always present on most indoor locations. This dataset has been collected in multiple testbeds, including large-scale and multifloor buildings, over the course of one year.

The platform provides some basic information of this dataset and a link to the original source of the data.[11]

### 4.3 magPIE Dataset

Magnetic field-based indoor positioning (Li et al., 2012, 2013; Montoliu et al., 2016) is an infrastructure-less approach which is based on the uniqueness of the disturbances in the magnetic field produced by the structural elements present in a scenario. The uniqueness of the disturbances can be used as a fingerprint, since it is stable over time.

*MagPIE* is a publicly available dataset for the evaluation of indoor positioning algorithms that use magnetic anomalies (Hanley et al., 2017). This dataset contains IMU and magnetometer measurements along with ground truth position measurements with

[11] http://ambiloc.org/.

centimeter-level accuracy. To produce this dataset, the authors collected over 13 hours of data from three different buildings, with sensors both handheld and mounted on a wheeled robot, in environments with and without changes in the placement of objects that affect magnetometer measurements.

The platform provides some basic information of this dataset and a link to the original source of the data.[12]

# 5 Methods Included in the Platform

Two different approaches are considered here for the fingerprinting-based location process: a deterministic, or nonparametric method; and a statistical, or parametric method. In the first, no statistical behavior is assumed, and the location problem is solved according to a set of observations whose positions are known; while the second method makes explicit use of distributions and statistical parameters of the data stored in the radio map to optimize the probabilities in the assignment of the estimated position.

## 5.1 Deterministic-Based Approach

The deterministic approach (Bahl and Padmanabhan, 2000; Marques et al., 2012; Torres-Sospedra et al., 2015a) relies on the well known $k$-Nearest Neighbors algorithm (*knn*) (Cover and Hart, 1967) to, given an RSSI vector, select the $k$ more similar training examples from the radio map. The similarity between the RSSI value vectors can be determined, for example, as the *Euclidean* distance between them, but other distance functions can be used instead (Torres-Sospedra et al., 2015a). Once the $k$ neighbors are selected, the method estimates the location of the user by calculating the weighted average of the neighbor's positions.

Fig. 7 shows a possible R source code of this method (with $k = 3$). The R dataframe *training.set* contains the RSSI values and the localization of the training points. The last two columns are the longitude (column name LONG) and latitude (column name LAT) of those points. *The validation.set* dataframe has the same structure. The complete description can be found at: http://indoorlocplatform.uji.es/methods/knn/.

## 5.2 Probabilistic-Based Approach

Given the limitations of sensors accuracy and the complex character of signal propagation, the RSSI vector stored for a particular position cannot have completely reliable and accurate information about the emitters signal strength. This uncertainty has been usually modeled by a normal distribution (Haeberlen et al., 2004), therefore many readings of the signals at the same position are needed to obtain a representative set of statistical parameters to model each RSSI present at that position. The more measurements for a particular location, the more reliable will be their inferred statistical parameters.

[12] http://bretl.csl.illinois.edu/magpie/.

```
k                <- 3
n_observations  <- nrow(training.set)
n_features      <- ncol(training.set) - 2
distances       <- matrix(0, 1, n_observations)
for (i in 1:n_observations) {
  distances[1, i] <- sqrt(sum((training_set[i, 1:n_features]
     - validation_set[1, 1:n_features])^2))
}
nearest       <- order(distances)[1:k]
weights       <- 1 / distances[nearest]
weights       <- weights / sum(weights)
est.longitude <- sum(training.set$LONG[nearest] * weights)
est.latitude  <- sum(training.set$LAT[nearest]  * weights)
```

**FIG. 7** A possible R source code for the deterministic method.

```
library(dplyr)
k           <- 3
delta       <- 10
data.means  <- training.set        %>%
               group_by(LONG, LAT) %>%
               summarise_each(funs(mean))
data.sds    <- training.set        %>%
               group_by(LONG, LAT) %>%
               summarise_each(funs(sd))
n_max       <- nrow(data.means)
n_waps      <- ncol(data.means) - 2
p           <- matrix(0, n_max, n_waps)
for (n in 1:n_max) {
    n_means <- as.numeric(data.means[n, 3:(n_waps + 2)])
    n_sds   <- as.numeric(data.sds[n, 3:(n_waps + 2)])

    for (j in 1:n_waps) {
      o1      <- validation.set[1, j] - delta
      o2      <- validation.set[1, j] + delta
      p1      <- pnorm(o1, mean=n_means[j], sd=n_sds[j],
         lower.tail=FALSE)
      p2      <- pnorm(o2, mean=n_means[j], sd=n_sds[j],
         lower.tail=FALSE)
      p[n, j] <- p1 - p2
    }
}
similarity    <- rowSums(p)
similar       <- order(similarity, decreasing=TRUE)[1:k]
weights       <- similarity[similar]/sum(similarity[similar])
est.longitude <- sum(data.means$LONG[similar] * weights)
est.latitude  <- sum(data.means$LAT[similar] * weights)
```

**FIG. 8** A possible R source code for the probabilistic-based method.

In the probabilistic-based approach (Haeberlen et al., 2004; Youssef and Agrawala, 2005; Madigan et al., 2005), the initial collection of RSSI observations associated with a particular point is transformed into a pair of vectors containing the means and the standard deviations of the RSSI for each beacon, and then the complete training data is stored as a set of statistical parameters. Then, given a test fingerprint, for each beacon, it is possible to estimate a probability value that expresses the similarity between the observation measurement at this beacon and the training data for a particular location. An evaluation of the total similarity for every location can be computed as a function of these individual probabilities.

The algorithm selects the $k$ training samples with higher probability and, similarly to the deterministic method, it estimates the location of the user by calculating the weighted average of the selected samples' positions.

Fig. 8 shows a possible R source code of this method (with $k = 3$). Input data has the same structure than in the deterministic method. The complete description can be found at:[13]

# 6 Experiments

The two methods explained in Section 5 have been tested with the four Wi-Fi-based datasets described in Section 4, using the tools included in the *Dashboard* section of the platform. Therefore, they are easily reproducible. All possible combination of the parameters has been tested. Only the combination of tuning parameters obtaining the best result is showed. In all cases, the test dataset has been used to assess the performance of the algorithms. The figure of merit used to provide an estimation of the performance of the methods is the mean localization error between the estimated position and the real one (internally known by the platform) of all test samples.

Table 5 shows the results obtained using the *UJIIndoorLoc* dataset. Note that there is no dataset for the building 0, floor 3 and for the building 2, floor 0, since there are no samples for these floors in the test set.

Table 6 shows the results on the *Tampere* dataset. In this case, only the results obtained with the deterministic approach are showed, since the probabilistic-based method can only be applied when there are enough samples at each position to calculate the estimation of the statistical parameters needed for the correct operation of this method.

Finally, Tables 7 and 8 show the results on the *IPIN2016 Tutorial* and *ALCALA2017 Tutorial* datasets. In both cases, all the samples are in the same building and floor, therefore it is not necessary to divide the data into subsets.

In the case of the *UJIIndoorLoc* dataset, the deterministic method provides better results than the probabilistic one in almost all the cases. The differences in the results obtained across buildings and floors depend on the quality of the radio map capture at each scenario and also on the structural characteristics of each scenario. According to the mean accuracy, the deterministic-based approach is preferable.

[13] http://indoorlocplatform.uji.es/methods/probabilistic/

**Table 5** Mean Positioning Error (in Meters) of Both Methods on the *UJIIndoorLoc* Dataset

| Building | Floor | Number of Samples | Deterministic | Probabilistic |
|---|---|---|---|---|
| 0 | 0 | 17 | 4.26 | 7.83 |
| 0 | 1 | 17 | 5.65 | 6.77 |
| 0 | 2 | 60 | 6.06 | 5.79 |
| 1 | 0 | 105 | 9.62 | 11.26 |
| 1 | 1 | 147 | 7.65 | 20.42 |
| 1 | 2 | 132 | 5.40 | 8.99 |
| 1 | 3 | 140 | 8.16 | 11.0 |
| 2 | 1 | 20 | 6.64 | 10.09 |
| 2 | 2 | 19 | 7.96 | 9.07 |
| 2 | 3 | 19 | 3.88 | 4.57 |
| 2 | 4 | 22 | 12.50 | 21.31 |
| | Mean | 704 | 7.18 | 10.64 |

**Table 6** Mean Positioning Error (in Meters) of Both Methods on the *Tampere* Dataset

| Building | Floor | Number of Samples | Deterministic |
|---|---|---|---|
| 1 | 1 | 156 | 9.83 |
| 1 | 2 | 110 | 14.21 |
| 1 | 3 | 118 | 8.01 |
| 1 | 4 | 105 | 13.03 |
| 2 | 1 | 61 | 15.87 |
| 2 | 2 | 77 | 8.38 |
| 2 | 3 | 37 | 6.74 |
| | Mean | 664 | 10.86 |

**Table 7** Mean Positioning Error (in Meters) of Both Methods on the *IPIN2016 Tutorial* Dataset

| Number of Samples | Deterministic | Probabilistic |
|---|---|---|
| 702 | 4.21 | 3.55 |

**Table 8** Mean Positioning Error (in Meters) of Both Methods on the *ALCALA2017 Tutorial* Dataset

| Number of Samples | Deterministic | Probabilistic |
|---|---|---|
| 405 | 5.03 | 2.53 |

**Table 9** Value of *RMID* Obtained for Each Dataset

| Dataset | Building | Floor | *RMID* |
|---|---|---|---|
| Alcalá2017 Tutorial | 1 | 1 | 0.05 |
| IPIN2016 Tutorial | 1 | 1 | 0.08 |
| UJIIndoorLoc | 0 | 1 | 0.22 |
| UJIIndoorLoc | 0 | 0 | 0.23 |
| UJIIndoorLoc | 2 | 2 | 0.29 |
| UJIIndoorLoc | 1 | 2 | 0.33 |
| UJIIndoorLoc | 2 | 3 | 0.33 |
| UJIIndoorLoc | 0 | 3 | 0.34 |
| UJIIndoorLoc | 2 | 1 | 0.35 |
| UJIIndoorLoc | 0 | 2 | 0.36 |
| UJIIndoorLoc | 2 | 0 | 0.49 |
| UJIIndoorLoc | 1 | 3 | 0.65 |
| UJIIndoorLoc | 1 | 1 | 0.66 |
| Tampere | 1 | 4 | 0.67 |
| UJIIndoorLoc | 1 | 0 | 0.67 |
| UJIIndoorLoc | 2 | 4 | 0.70 |
| Tampere | 1 | 3 | 0.72 |
| Tampere | 1 | 2 | 0.77 |
| Tampere | 1 | 1 | 0.80 |
| Tampere | 2 | 1 | 0.89 |
| Tampere | 2 | 2 | 0.90 |
| Tampere | 2 | 3 | 0.91 |

The data is ordered by ascending *RMID* value.

There is also a high variability across buildings and floors in the results obtained for the *Tampere* dataset due to the same reasons than in the *UJIIndoorLoc* dataset.

In the case of the *IPIN2016 Tutorial* and *ALCALA2017 Tutorial*, the results are very similar and in both cases the probabilistic approach is preferable. Note that, in the *ALCALA2017 Tutorial* dataset, the difference is quite significant since the probabilistic-based approach can deal with the unintentional mistakes introduced by some of the dataset creators. Results obtained for these datasets are better than the ones obtained for the *UJIIndoorLoc* and the *Tampere* datasets since the scenarios of the Tutorial datasets correspond to small areas and more fingerprints per $m^2$ than the other two. Therefore position error is lower.

Table 9 shows the Radio Map Inherent Difficulty (RMID) value of each dataset (Sansano et al., 2017). This measure gives an estimation of the inherent difficulty of a radio map to obtain accurate estimates. According to this value, *IPIN2016 Tutorial* and *ALCALA2017 Tutorial* are scenarios where it is easier to obtain accurate results. However, as confirmed by the results showed in Tables 5 and 6, the RMID value of the *UJIIndoorLoc* and *Tampere* datasets shows that both datasets are quite complex and therefore, it is quite difficult to obtain positioning estimation with high accuracy.

Taking into account all the datasets, the mean localization error of the deterministic-based method is 6.79 m without including the results of the *Tampere* dataset, and 8.21 including it. The mean error of the probabilistic-based method is 9.47. Therefore, according to the results, in general, it seems that the deterministic-based method is preferable. However, taking into account the differences of the four scenarios, the deterministic-based approach gets better results in big scenarios with low density of data (*UJIIndoorLoc* and *Tampere* datasets), while the probabilistic based one is preferable in small ones with high density of data (*IPIN2016 Tutorial* and *ALCALA2017 Tutorial* datasets).

Note that the results obtained with the methods included in the platform can be effectively improved using more sophisticated algorithms, and also using modern machine learning techniques. For instance, the ranking of the *UJIIndoorLoc* dataset shows better results than the ones presented in Table 5, since they are the best results obtained in the *Wi-Fi fingerprinting in large environments* (IPIN'15) competition.

# 7 The Platform in Use

The performance of the platform was tested during the 2017 Fingerprinting-based Indoor Positioning tutorial held in the School of Engineering of the University of Alcalá. As an activity of the course, an indoor localization competition took place using the *ALCALA2017 Tutorial* dataset.

The 15 attendees were invited to use the platform to upload the results of their proposals. Some of them used the *Dashboard* section of the platform to test different parameter configurations of the localization methods included in the platform, and others manually programmed their own method from the source code provided by the course instructors. After a very competitive and exciting competition, the winner team got an error of only 2.14 m using a probabilistic method. This result is even better than the best one that can be directly obtained using the Dashboard included in the platform.

In general, tutorial attendees were able to easily use the platform, mainly the *Dataset* download section, the *Dashboard* section and, obviously, the *Ranking* sections. Almost no queries to the course introduction were produced, showing the effective user-centered design applied to the platform.

# 8 Conclusions

In this chapter, the *IndoorLoc Platform* has been presented. It is a public repository for comparing and evaluating indoor positioning algorithms. The proposed web platform can be used to download datasets, learn how some well-known algorithms work, study the source code of those algorithms, test the methods, and even upload results of the user's methods to check the accuracy when comparing against the results provided by other methods already included in a ranking, among other functionalities.

To present a real example of the usage of the platform, a comparative study of the accuracy of two well-known fingerprinting-based indoor localization algorithms, using four of the datasets included in the platform, have also been presented. According to the results obtained, the deterministic-based approach gets better results in big scenarios while the probabilistic based one is preferable in small scenarios. These experiments are easily reproducible using the tools included in the platform.

This web platform is an ongoing project, and future versions will implement new algorithms and include more datasets, with the aim to provide an interesting tool for researchers and become a reference web platform for indoor positioning research. For this purpose, researchers are invited to include more methods and datasets in the platform.

# Acknowledgments

The authors gratefully acknowledge funding from the Spanish Ministry of Economy and Competitiveness in the projects: "Proyectos I+D Excelencia" TIN2015-70202-P and "Redes de Excelencia" TEC2015-71426-REDT.

# References

Bahl, P., Padmanabhan, V.N., 2000. RADAR: an in-building RF-based user location and tracking system. In: Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00), pp. 775–784.

Barsocchi, P., Crivello, A., Rosa, D.L., Palumbo, F., 2016. A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting. In: Proceedings of the Seventh Conference on Indoor Positioning and Indoor Navigation (IPIN'16).

Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 13 (1), 21–27.

Cramariuc, A., Lohan, E.S., 2016. Open-access WiFi measurement data and python-based data analysis. Available from: http://www.cs.tut.fi/tlt/pos/meas.htm.

García, S., Molina, D., Lozano, M., Herrera, M.F., 2009. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the cec'2005 special session on real parameter optimization. J. Heuristics 15 (6), 617–644.

Goldbloom, A., Hamner, B., Moser, J., Cukierski, M., 2017. Kaggle: your home for data science. Available from: https://www.kaggle.com/.

Haeberlen, A., Flannery, E., Ladd, A.M., Rudys, A., Wallach, D.S., Kavraki, L.E., 2004. Practical robust localization over large-scale 802.11 wireless networks. In: Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (ModiCom'04), pp. 70–84.

Han, D., Jung, S.H., Lee, M., Yoon, G.-W., 2014. Building a practical Wi-Fi-based indoor navigation system. IEEE Pervasive Comput. 13, 72–79.

Hanley, D., Faustino, A.B., Zelman, S.D., Degenhardt, D.A., Bretl, T., 2017. Magpie: a dataset for positioning with magnetic anomalies. In: Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN'17).

He, S., Chan, S.G., 2016. Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. IEEE Commun. Surv. Tutorials 18 (3), 466–490.

Lemic, F., Handziski, V., Wirstrom, N., Van Haute, T., De Poorter, E., Voigt, T., Wolisz, A., 2015. Web-based platform for evaluation of RF-based indoor localization algorithms. In: Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW'15).

Li, B., Gallagher, T., Dempster, A.G., Rizos, C., 2012. How feasible is the use of magnetic field alone for indoor positioning? In: 3th International Conference on Indoor Positioning and Indoor Navigation.

Li, B., Gallagher, T., Rizos, C., Dempster, A., 2013. Using geomagnetic field for indoor positioning. In: Proceedings of the International Global Navigation Satellite Systems (IGNSS) Society Symposium.

Lichman, M., 2013. UCI machine learning repository. Available from: http://archive.ics.uci.edu/ml.

Lymberopoulos, D., Choudhury, R.R., Yang, X., Sen, S., 2014. Microsoft indoor localization competition (IPSN'14). Available from: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2014.

Lymberopoulos, D., Liu, J., Yang, X., Naguib, A., Rowe, A., Trigoni, N., Moayeri, N., 2015. Microsoft indoor localization competition (IPSN'15). Available from: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2015.

Lymberopoulos, D., Liu, J., Zhang, Y., Dutta, P., Yang, X., Rowe, A., 2016. Microsoft indoor localization competition (IPSN 2016). Available from: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2016.

Lymberopoulos, D., Liu, J., Bocca, M., Sequeira, V., Trigoni, N., Yang, X., 2017. Microsoft indoor localization competition (IPSN 2017). Available from: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2017.

Madigan, D., Einahrawy, E., Martin, R.P., Ju, W.-H., Krishnan, P., Krishnakumar, A.S., 2005. Bayesian indoor positioning systems. In: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05), pp. 1217–1227.

Marques, N., Meneses, F., Moreira, A., 2012. Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning. In: Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN'12), pp. 1–9.

Moayeri, N., Ergin, O., Lemic, F., Handziski, V., Wolisz, A., 2016. Perfloc: an extensive data repository for development and a web-based capability for performance evaluation of smartphone indoor localization apps. In: Proceedings of the 27th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'16).

Montoliu, R., Torres-Sospedra, J., Belmonte, O., 2016. Magnetic field based indoor positioning using the bag of words paradigm. In: Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN'16).

Montoliu, R., Sansano, E., Torres-Sospedra, J., Belmonte, O., 2017. IndoorLoc platform: a public repository for comparing and evaluating indoor positioning systems. In: Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN'17).

Nahrstedt, K., Vu, L., 2012. CRAWDAD dataset uiuc/uim (v. 2012-01-24). Available from: http://crawdad.org/uiuc/uim/20120124.

Popleteev, A., 2017. AmbiLoc: a year-long dataset of FM, TV and GSM fingerprints for ambient indoor localization. In: Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN'17).

Potortì, F., Barsocchi, P., Girolami, M., Torres-Sospedra, J., Montoliu, R., 2015. Evaluating indoor localization solutions in large environments through competitive benchmarking: the EvAAL-ETRI competition. In: Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation (IPIN'15).

Sansano, E., Montoliu, R., Torres-Sospedra, J., 2017. A novel methodology to estimate a measurement of the inherent difficulty of an indoor localization radio map. In: Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN'17).

Talvitie, J., Lohan, E.S., Renfors, M., 2014. The effect of coverage gaps and measurement inaccuracies in fingerprinting based indoor localization. In: Proceedings of International Conference on Localization and GNSS 2014 (ICL-GNSS'14).

Torres-Sospedra, J., Montoliu, R., Usó, A.M., Avariento, J.P., Arnau, T.J., Benedito-Bordonau, M., Huerta, J., 2014. UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In: Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN'14), pp. 261–270.

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, O., Huerta, J., 2015a. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Syst. Appl. 42 (23), 9263–9278.

Torres-Sospedra, J., Rambla, D., Montoliu, R., Belmonte, O., Huerta, J., 2015b. UJIIndoorLoc-Mag: a new database for magnetic field-based localization problems. In: Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation (IPIN'15).

Torres-Sospedra, J., Jiménez, A.R., Knauth, S., Moreira, A., Beer, Y., Fetzer, T., Ta, V.-C., Montoliu, R., Seco, F., Mendoza-Silva, G.M., Belmonte, O., Koukofikis, A., Nicolau, M.J., Costa, A., Meneses, F., Ebner, F., Deinzer, F., Vaufreydaz, D., Dao, T.-K., Castelli, E., 2017a. The smartphone-based offline indoor location competition at IPIN 2016: analysis and future work. Sensors 17 (3).

Torres-Sospedra, J., Moreira, A., Knauth, S., Berkvens, R., Montoliu-Cols, R., Belmonte-Fernndez, O., Trilles, S., Nicolau, M.J., Meneses, F., Costa, A., Koukofikis, A., Weyn, M., Peremans, H., 2017b. Realistic evaluation of indoor positioning systems based on Wi-Fi fingerprinting: the 2015 EvAAL-ETRI competition. J. Ambient Intell. Smart Environ. 9, 263–279.

Wu, C., Yang, Z., Liu, Y., Xi, W., 2013. WILL: Wireless indoor localization without site survey. IEEE Trans. Parallel Distrib. Syst. 24 (4), 839–848.

Youssef, M., Agrawala, A., 2005. The Horus WLAN location determination system. In: Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys'05), pp. 205–218.