Chapter 3
Foundations

Mobile IR can be quite easily mistaken as "squeezing IR on a small screen". Actually the situation is more complex, and there are some conceptual and foundational aspects that deserve to be properly addressed. In this chapter we start with more general and introductory material about the history of mobile devices, their technological development, and some social aspects in Sects. 3.1 and 3.2. Then we turn to more technical details. Section 3.3 describes the studies on how the concept of relevance changes in the mobile world. Section 3.4 presents the research on IR models adapted to the new scenario.

3.1 The Mobile Phone Device

The history of mobile devices, although rather short, is probably too rich to be analyzed in detail here, and it is well described elsewhere, including detailed Wikipedia pages. We briefly summarize the most important concepts only, and focus on what will be useful in the following.

The first mobile phones were demonstrated in the 70s and released in the 80s, but a large scale adoption started in the 90s. The first models suffered from several limitations: very small and low resolution screens, cumbersome input/output capabilities, low bandwidth, minimal computational power, and short battery life. This has constantly changed, with increasing higher resolution screens, more convenient input/output by means of better hardware and software keyboards and touch screens, wider bandwidth and capability to connect to Wi-Fi networks, more computational power, and improved battery life.

F. Crestani et al., *Mobile Information Retrieval*,
SpringerBriefs in Computer Science, DOI 10.1007/978-3-319-60777-1_3

Of course the history is not without unsuccessful attempts; for example, the first attempts of web enabled phones, based on WAP (Wireless Application Protocol), have had a rather short life. And in the first years it was clear that the technological limitations were just temporary. For these reasons, several early experiments in Mobile IR, and in mobile computing in general, have been performed using PDA (Personal Digital Assistants), that did not suffer from the limitations of earlier mobile phones — we will see examples in the following.

Indeed, the technological development has been impressive: today the screens have a resolution higher than the resolution power of the human eye; the sequence of network protocols with increasing bandwidth (GPRS, EDGE, 3G, 4G, LTE to name a few), allows extremely high data transfer rate when the network coverage is adequate; battery life is rarely an issue; and, concerning computational power, as it is often stated, today everyone can have in his own pocket more computational power than what could fit in a room just some decades ago.

To focus on the recent ten years or so, one important turn point has been the launch of the *iPhone*, commercialized in 2007. Leaving aside the advertising slogans presenting it as a combination of music device, revolutionary phone, and internet communicator featuring a touch screen, the iPhone has changed the way people use their phones, and made possible more advanced, sophisticated, and powerful interaction modalities, as well as novel services.

Another important aspect is that mobile devices are truly *multimedia* devices: the availability of high resolution cameras, sound input/output devices, and video has increased the amount of multimedia material to be searched (sometimes tagged on the basis of the context).

It is also important to understand that the mobile devices are equipped with several other *sensors*, like GPS and location detection, direction, movement, temperature, touch, light, noise, etc. The importance of this aspect comes from the fact that the mobile device is truly "in the real world", and this has important consequences. For example, being in the physical world enables location-based (or, more generally, context-based) search, that can be raised to a level unreachable with classical desktop devices. Also, the interplay among all these sensors is quite complex [28], and perhaps not yet understood; for example, of course location can enhance image recognition when trying to understand where a photo has been taken and what is its subject; but conversely the location can be derived by a photo taken with the device, and/or maybe its tags. A further and related aspect to be mentioned is that the mobile device is not an isolated device, and always less so; rather, there is a clear trend towards an increasing *integration* with both desktop computers and other devices and sensors like, e.g., the "glasses" or the armbands capable of monitoring motion, gesture, and all kind of lifestyle, fitness, and health data. This wide array of sensors poses challenges and offers opportunities; probably the potential is still unexplored and will be an important research topic in the next years.

Also, the meaning of the term "mobile device" itself has changed over the years: from mobile phone, through PDA, and today including *tablets* with much improved input output capabilities by means of larger touch screens.

Finally, besides being in a "physical environment", the device and its owner are also in a "social environment", in several respects. Mobile devices are used extensively to access social networks (e.g., Facebook, Twitter, etc.), sometimes with a strong location attention (e.g., Foursquare). The first studies are being published that show how to infer the current activity of the user from sensors data. It is theoretically possible to understand whom the user is with.

In the following we will see many concrete examples of how the mobile device features described here cause deep differences between IR and Mobile IR. For example, in Sect. 4.2 we discuss how the availability of cameras and sound input/output devices makes multimedia collections a natural environment for Mobile IR, where photos and videos are created and searched; in Chap. 6 we analyze more in detail the effects of size, screen, and keyboard on a new kind of user interface; in Sect. 4.4 we discuss how being in the real world leads to searching a new kind of collection, consisting in Points of Interest (POIs); and in Chap. 7 we emphasize how the many sensors can be exploited for location- and context-aware search.

3.2 The Role of Technology and Society

When looking at its brief history, it is clear that Mobile IR is a technology grounded field: Mobile IR exists because the mobile devices exist, and the field has followed the technological development. Indeed, in the case of mobile devices, the technological development is so fast paced that sometimes commercial products seem even ahead of the results of scientific research. This is perhaps already true to some extent for Web search, but it is even more important for Mobile IR: one paradigmatic example is the introduction of the iPhone that has changed the way people use their mobile phones. The device, however, although crucial is not the only element, as the services that have been and are being proposed change the way people use the device as well; think, for example, of SIRI, Google Now, Foursquare, etc. Technology must be interpreted in a broad sense. It is also likely that future development will depend on new technologies as well.

Moreover, it would be shortsighted to consider Mobile IR simply as "adapting IR to the latest fancy mobile phone", let alone "squeezing the IR user interface to a small screen and keyboard". As already mentioned, the development is not only on the device, but on networks, bandwidth, software, and services. Therefore, to fully understand the field it is important to take into account the latest relevant technological developments on devices, products, and services.

Finally, the technological development is related to changes in the society and in human habits. Although the first handheld mobile phones appeared in the 80s, there has been a constant growing adoption rate since the 90s that has led to today situation, where there are probably as many active mobile phones on the Earth as human beings. This has direct consequences, like the fact that, since 2015 there are more queries to search engines by mobile devices than by desktop computers.

3.3 The Concept of Relevance

Relevance is a crucial notion for IR, and it has been studied extensively in the IR literature. Some attempts have been made to study at a conceptual level this foundational notion from a Mobile IR standpoint, highlighting the differences between the classical and the mobile environment.

Some features of relevance are examined by Coppola et al. [42], that extends to the mobile world Mizzaro's framework of relevance [131]. Mizzaro proposes four dimensions to characterize the various types of relevance:

- information resources,
- user problem,
- time, and
- components.

The first dimension models what the user is searching for, and makes explicit the fact that relevance can deal with the information resources at three levels of abstraction: a document containing information; a surrogate of the document (e.g., title, keywords, and abstract); and the information received by the user, as she perceives it. The user problem (second dimension) has four levels of abstraction: her real information need; the information need as she perceived it; the information need as she can express it in natural language; and the query as it is expressed in the system language. The third dimension 'time' refers to the steps in which the information flows with the interaction from the moment in which the user's real information need arises to the moment in which it is satisfied. The fourth dimension 'components' lists the aspects that compose the first two dimensions: the topic the user is interested in; the task or activity she aims to perform; and the context in which everything happens.

Coppola et al. [42] modify and extend Mizzaro's framework. One of the changes is that, in the first dimension, one further level of abstraction is added, thus getting four resources: the actual entity (or thing). In the mobile world, relevance addresses real/physical world entities and not only documents. A related change is on the second dimension, where a corresponding 'thing need' is added. These two changes depend on the increased importance of the real world (which is not surprising at all, being the device "in the physical world", see Sect. 3.1), and the same motivation underlies two further remarks made by the authors: the increased importance of time (related to the third dimension) and of context (fourth dimension). Another observation related to context is that, in the Mobile IR scenario, context is both more complex and more easily derivable than in the classical IR case. This position, and the examples presented in the second part of the paper, inspired the future work by the same research group [43–46, 132] (discussed at length, with similar approaches, in the following chapters).

The same concept of relevance is discussed (under the name of Geographical relevance) in another conceptual paper by Raper [152], specifically from a background, and a standpoint, in geography. Some considerations are similar: information needs, named geographic information needs, are discussed; geographic information objects are studied; geographic relevance is defined on this basis; time is often mentioned; a detailed model of geographic relevance is proposed; etc. However there are also some differences, in part due to the different background (geography): a taxonomy of geographic information needs is proposed along two dimensions (intentions vs. tasks and geo-representation vs. geo-context); it is remarked that, as it is customary in geography, a model of geographic space can be defined using either geo-centric coordinates or places and landmarks; space is often mentioned together with time, suggesting that space/time might be the proper concept to take into account; and more in general, whereas Coppola et al. [42] — implicitly — assumed that the various entities in the model are not influencing each other, in Raper's paper the need to understand such influences is explicitly mentioned. One example in this respect is the space/time concept. Another example is discussed specifically by Reichenbacher and De Sabbata [153], who remark the importance of the spatial organisation of geographic information objects: they have a "spatial layout [that] can have historical, geographical, economical, or social reasons" [153, p. 69] (like, for example, objects in a cluster or hierarchically organized), and that should be taken into account when determining their relevance.

The last paper on this topic is by Sabbata et al. [158], that build on the papers by Mizzaro [42, 131] and further extend their models along the following four directions.

- In the first dimension, it is specified that sometimes there is no document at all and the system internal representation of the sought item is a descriptor of a thing in the real world, rather than a descriptor of a document that refers to the thing.
- The time third dimension is modified: the basic observation is that as the user is moving, the time needed to reach a specific point in the surroundings will change as time goes on, maybe even rapidly. This calls for an integrated space-time dimension that extends the time one.
- It is noted that there are different representations of the world, including that perceived by the user, that represented in the system, that documented in stored information, and the real one. This is accounted for by proposing to add a further world dimension to the framework.
- A different components fourth dimension is proposed, that includes topic, activity (which replaces task and is further analyzed in more detail), user's preferences, social aspects, mobility, and context.

One important take home message that we can draw is that being the device "in the physical world" (Sect. 3.1) makes also real world entities searchable. Another important lesson is that relevance in Mobile IR is truly multidimensional, and beyond-topical features like time and location must be taken into account, probably more than in the classical desktop IR case.

A more practical study of relevance in Mobile IR is by Verma and Yilmaz [191], who compare relevance judgments on desktop computers and mobile devices. Results of a crowdsourcing experiment show that gathered relevance labels differ in the two cases, whereas inter-rater agreement is similar. It also is found that mobile judges are faster.

3.4 Mobile IR Models

The above observations on the nature of mobile relevance are the starting point of more pragmatic work like Bouidghaghen et al. [19], where a model is proposed to deal with combinations of relevance criteria. Although the approach could be more general, it is particularly suited to Mobile IR. The relevance criteria included in Bouidghaghen et al.'s model are topic, user's interest, and user's location. Each of them receives its own value: topic is measured using BM25, interest by a tf.idf schema, and location by a geographical weighting function based on the frequency (a sort of tf) of both location and sub-locations in the document. The three values are then combined by means of "prioritised aggregation operators". These operators are defined on the basis of user's preference, expressed as a rank, on the criteria, but they do not simply compute some weighted average; rather, their definition is based on: (i) the preference rank of the criteria, (ii) the weight of the previous more important criterion, and (iii) the satisfaction degree of the document with respect to the previous more important criterion. An experimental evaluation on an in-house built collection shows the effectiveness of the approach.

The issue of how to exploit in local search new features beyond the textual ones used in Web search is discussed also by Berberich et al. [14]. They derive distance and popularity features from external sources and show that such features are effective in improving effectiveness.

Another technical contribution is by O'Hare and Murdock [144], who study the problem of location identification from arbitrary text. To this aim, they exploit a core IR technique and develop statistical language models for locations, that are defined as square cells of 1 by 1 km, 10 by 10 km, and 100 by 100 km. That is, each location (square cell) is considered a document, and its terms are the Flickr textual tags of the photos geotagged in that area. The term distribution defines the language model of that location. The models are then tested to predict the location of (other) geotagged Flickr photos by using as queries only their textual tags, and on another dataset (CoPhIR) as well. Several classical language model approaches are adapted and used, like maximum likelihood estimate, prior location probability (to take into account location popularity), and smoothing (to refine the term distributions of a location on the basis of its neighbourhoods). The experimental evaluations show that reasonable accuracy can be obtained at various granularity levels. For example, 1 by 1 Km locations can be predicted with a 17% accuracy, and 3 by 3 Km ones with 40%. The experimental results also confirm that smoothing can be effective.

Other researchers focus on assigning positions to queries, and exploiting the position to improve retrieval. Although the relevance to Mobile IR of this research issue is perhaps more marginal, and it can be useful for search in general (indeed the term "mobile" does not even occur in some of the papers), it deserves anyway a short mention at least. Backstrom et al. [11] develop a probabilistic model to assign to each query (they work on the complete Yahoo! query logs) a center plus a spatial dispersion. Yi et al. [203] aim at discovering implicit user's local intent in queries. They develop a classifier able to effectively categorize queries at the city level. A similar approach is followed by Lu et al. [122] who focus on implicit local intent queries, i.e., queries that are clearly searching for things in a particular location but without any explicit location mention. They develop a probabilistic based model to automatically classify the queries that have an implicit intent. An attempt to improve retrieval effectiveness of those queries is made by first inferring user's location from the IP address and then either (i) expanding the query by adding the location, or (ii) re-ranking the retrieved documents on the basis of the location. In an experimental evaluation, both approaches improve retrieval effectiveness, with re-ranking being significantly more effective.

3.5 Conclusions

After a brief mention of general issues related to devices, technology, and society, this chapter has discussed how the notion of relevance needs to be adapted to the Mobile IR world, highlighting some important changes. We will see consequences of these conceptual approaches in the following (for example, in Sect. 4.4, and in Chaps. 7 and 8). We can say that the discussed papers on relevance contain cues, hints, and examples that, although at a conceptual level, are perhaps still to be fully understood and exploited. In the last section we have discussed how Mobile IR affects the foundational topic of IR models, and presented a sample of the Mobile IR models that have been proposed.

http://www.springer.com/978-3-319-60776-4

Mobile Information Retrieval
Crestani, F.; Mizzaro, S.; Scagnetto, I.
2017, VI, 110 p. 9 illus., Softcover
ISBN: 978-3-319-60776-4