Identity Crisis of Ubicomp? Mapping 15 Years of the Field's Development and Paradigm Change

Yong Liu, Jorge Goncalves, Denzil Ferreira, Simo Hosio, Vassilis Kostakos
Department of Computer Science and Engineering, University of Oulu
Erkki Koiso-Kanttilan katu 3, FI-90014 Oulu, Finland
Firstname.lastname@ee.oulu.fi

ABSTRACT

The rapid growth of the Ubicomp field has recently raised concerns regarding its identity. These concerns have been compounded by the fact that there exists a lack of empirical evidence on how the field has evolved until today. In this study we applied co-word analysis to examine the status of Ubicomp research. We constructed the intellectual map of the field as reflected by 6858 keywords extracted from 1636 papers published in the HUC, UbiComp and Pervasive conferences during 1999-2013. Based on the results of a correspondence analysis we identify two major periods in the whole corpus: 1999-2007 and 2008-2013. We then examine the evolution of the field by applying graph theory and social network analysis methods to each period. We found that Ubicomp is increasingly focusing on mobile devices, and has in fact become more cohesive in the past 15 years. Our findings refute the assertion that Ubicomp research is now suffering an identity crisis.

Author Keywords

Ubicomp; identity crisis; ubiquitous computing, social network analysis, bibliometric; paradigm change

ACM Classification Keywords

K.2 History of Computing

INTRODUCTION

In the 2012 UbiComp conference paper "*What next, ubicomp? Celebrating an intellectual disappearing act*" [1], Abowd provided a thoughtful assessment of the Ubicomp field's maturity, and highlighted an identity challenge it faces. This spurred constructive debate amongst researchers in the field. However, much of this broader debate has been driven by individual researchers' tacit understanding of where the field currently stands, and may be biased by a number of different factors.

Here, we attempt to mitigate to some extent the subjectivity of such an assessment. By analyzing 1636 papers published in the 15 years of the UbiComp/Pervasive/HUC conference series we hope to provide a more objective and clear understanding of the progression of the field, where it currently stands, and which research agendas will be potential driving forces in the field in years to come. In response to the identity crisis debate, we analyze whether the field has become increasingly cohesive or fragmented over time. We also reflect on the impact of the changes of research topics in the field. For instance, we consider whether the evolution of research topics in UbiComp should be regarded as evidence of an identity crisis or a natural metabolism in response to the changes of external technological environment.

Our analysis uses techniques from hierarchical cluster analysis and graph theory, through the use of co-word analysis artifacts such as strategic diagrams and graphs. Co-word analysis is part of the co-occurrence analysis methods. It is a widely-applied bibliometric approach to describe the interactions among concepts, ideas, and problems and to explore the concept network within a scientific area [7,8]. Co-word analysis rests on the assumption that a paper's keywords constitute an adequate description of its content as well as the links the paper established between problems: two keywords co-occurring within the same paper are an indication of a link between the topics to which they refer to [9]. The presence of many co-occurrences around the same word, or pair of words, suggests a locus of strategic alliance within articles that may reflect a research theme [9].

More importantly, by measuring the association strength of keywords produced in a specific scientific discipline, co-word analysis allows researchers to identify key patterns and trends within a scientific discipline [19,28,40]. It is assumed that a specific keyword with adequate frequency refers to a particular research topic while a cluster or pattern of keywords refers to a specific research direction or research theme. A change in a research theme (*e.g.*, declining or emerging research interest) or a change of topics within a research theme implies a paradigm change.

Differing from conventional co-word studies, in this paper we developed a supervised algorithm to extract keywords from the abstract of the papers. This way we attempt to

UbiComp '14, September 13 - 17, 2014, Seattle, WA, USA

ACM 978-1-4503-2968-2/14/09…$15.00.
http://dx.doi.org/10.1145/2632048.2632086

reduce the inconsistencies caused by authors' subjective indexing of keywords throughout the history of the conference.

RELATED WORK

The main concepts we use in our analysis are *keywords*, *networks*, and *clusters*. *Keywords* are associated with each paper, and two keywords associated with the same paper are linked to form a *network* (or graph) of keywords. Analysis of this network helps us identify *clusters* (a set of closely-related keywords).

Our co-word analysis reduces a large space of descriptors (*i.e.*, keywords) into a network graph (*i.e.*, multiple related smaller spaces). Easier to comprehend but still retaining crucial information, this approach visualizes the interrelated concepts [14] and intellectual structure of a discipline into a map of the conceptual space of this field, and a time-series of such maps produces a trace of the changes in this conceptual space [19]. Co-word analysis has been widely utilized in mapping the conceptual networks of a diversity of disciplines, like business intelligence [52], consumer behavior [43], software engineering [14], patent analyses [11], biology [2,9], education [50], human computer interaction [41] and library and information science [19]. As such, it makes sense to apply this technique to enrich our understanding of the flagship ubiquitous computing conference.

Given a network of keywords, we can use network analysis and strategic diagrams to characterize the field. Keywords and clusters have different properties, depending on how they are linked. For instance, bridges between two nodes (*i.e.*, linked nodes) in a network perform a valuable function in allowing communication and facilitating the flow between otherwise isolated regions of the network, also known as *structural holes* [46]. The greater the number of bridges associated with a research topic or theme, the more it serves to connect otherwise isolated research topics or themes. Keywords with a great number of structural holes serve as the *backbone* of the whole network. If these are removed from the network, the whole network will collapse into a number of separated and unconnected research sub-fields, therefore losing its scientific cohesion and identity.

When computing a network's *core-periphery* structure, it becomes possible to determine which nodes are part of a densely connected core (*i.e.*, with a higher number of bridges) and which are part of a sparsely connected periphery [6,51]. Core nodes are typically well connected to peripheral nodes. Peripheral nodes are sparingly connected to a core or to each other. In a keyword network it is expected that, as the body of knowledge grows, peripheral nodes become core nodes, thus allowing for the emergence of new peripheral nodes. Research topics with a high core value delimit the main body of Ubicomp knowledge, and represent important knowledge-growing points of the main body of the field.

In our work we rely on two graph theory concepts to map the field of Ubicomp: *density* and *centrality*, defined as follows:

- *Density*, or internal cohesion, measures the *strength of the links that tie together the cluster of keywords* making up the research theme. This can be understood as a measure of the theme's development [26,43]. Density offers a good representation of the cluster's capacity to maintain itself and to develop over the course of the time in the field [8,26]. The higher the density, the more coherent the cluster is and the more likely it is to contain inseparable expressions;
- *Centrality* measures the *degree of interaction of a theme with other parts* of the network [46]. In other words, it measures the strength of external ties of a research theme to other research themes, and can be referred to as a measure of the importance of a theme in the development of the entire research field [43]. The greater the number and the strength of a theme's connections with other themes, the more central this theme will be to the whole network [5].

The measurements of group density and centrality are based on the keyword co-occurrence matrix. The centrality is calculated based on K-step=1.

By combining both concepts we then created a *strategic diagram*. Strategic diagrams are two-dimensional plots that have been widely used in prior co-word analysis studies [8,14,40,43]. The x-axis shows *the strength of interaction between a specific research theme with others* (*i.e.*, centrality). The y-axis reflects the density of the research theme, or the *internal cohesion of a specific research theme* (see Figure 1).

The location of a given research theme within this strategic diagram characterizes the theme in the context of the whole discipline:

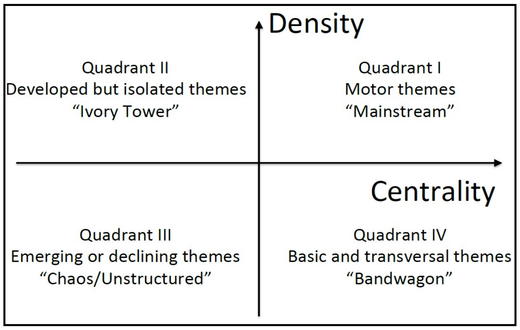


Figure 1. Strategic diagrams use density and centrality to classify research themes.

- *Quadrant I* (Figure 1, top-right): both internally coherent and central to the research network in question. Known as the motor-themes of the discipline given that they present strong centrality and high density;

- *Quadrant II* (Figure 1, top-left): coherent but low centrality themes. These themes are internally well structured and indicate that a constituted social group is active in them. However, they have rather unimportant external ties resulting in specialized work that is rather peripheral to the work being carried out in the global research network;
- *Quadrant III* (Figure 1, bottom-left): weakly developed with marginal interest in the global research network. These themes have low density and low centrality, mainly representing either emerging or disappearing themes;
- *Quadrant IV* (Figure 1, bottom-right): weakly structured themes. These are strongly linked to specific research interests throughout the network but are only weakly linked together. Hence, prior work in these themes is under-developed yet transversal, with potential to be of considerable significance to the entire research network.

METHOD

There is a substantial inconsistency in the use of author keywords throughout the history of the three conferences we surveyed (HUC, Ubicomp, Pervasive). We decided to ignore those and focus on the papers' abstracts. Using a concordance analysis tool (CasualConc) we identified sequences of words that appeared frequently in the whole corpus (*i.e.*, the union of all abstracts). This approach can identify both single-word and multi-word keywords. Long clusters of keywords that were manually identified by the researcher as "generic" statements were removed. Such statements include the phrases "in this paper we present", "in this paper we argue", and "we present a novel".

The remaining raw text of the abstracts was processed by removing all stop words ("and", "or", "of"), and then we used Porter's stemming algorithm [47] on the remaining text. This algorithm removes the ending of words and retains the "stem". This is done so that words like "technological", "technologies", "technology" are reduced to their stem "technolog." The stemmed corpus was again subjected to concordance analysis to identify the most popular clusters of words. Our process began with the longest clusters (*i.e.*, those clusters with most words) and working down the list to clusters containing only 2 words.

For each cluster we chose one of two outcomes: i) remove, or ii) convert to a keyword token. Many clusters were removed at this stage because they were judged to be "generic" statements. We note that these clusters were not picked up in the initial step because they could exist in many variations when considering the removed word endings and stop words. For instance, the stemmed cluster "paper describ design" could be expanded into multiple versions of full text: "paper describes the design", "paper describes designing", "paper we describe how we designed", and so on. Clusters that were perceived to be non-generic were converted to a keyword token. Different clusters could be converted to the same keyword token. For instance, the clusters "field study" and "field trial" were both converted to the keyword "field trial." The concordance analysis was iterated every time a new keyword token was created. We repeated this process until every cluster was either removed or converted to a keyword token.

This process is effectively a bottom-up semi-automated approach to building a dictionary that describes all words of "value" in the original abstracts. This was far from an objective or automated process, and substantial human input and knowledge went into the creation of the dictionary. For example, the following clusters were all converted to the keyword token "qualitative methods": experi sampl, diary studi, focus group, self report, interview, survey, ethnograph, questionnair. Similarly, all the following clusters were converted to the keyword token "sensing": sensor network, sensor data, sens system, sensor node, monitor system, sens data, sens devic, sensor base, sensor devic, sensor, acceleromet.

Lastly, an important decision we had to make was how many keywords tokens to generate to describe the intellectual field of Ubicomp. A previous analysis of the CHI conference proceedings between 1994-2013 (20 years, 3152 articles) showed that less than 100 keywords were enough to describe that intellectual field [41]. In our case we generated 67 keywords for the whole corpus of 1372 articles.

DATA

We collected data on all papers published in the HUC, UbiComp and Pervasive conferences since 1999. The two latter conferences are originally derived from the handheld and ubiquitous computing (HUC) conference that was held in 1999 and 2000. In 2013 the two conferences merged. 1372 papers were identified through DBLP, which includes all the "archived" proceedings between 1999 and 2013 with DOI numbers. The abstract of every article listed on DBLP was retrieved from the original publishers. All abstracts were in English.

With our analysis, we extracted 67 unique keywords from the abstracts. We note that "ubiquitous computing" is an inherent focus of the conference, therefore cannot provide interesting information to the analysis [29,39]. Therefore, the keyword token derived from phrases including 'ubiquitous computing' and 'pervasive computing' was removed from the analysis, resulting in a final sample of 66 different keywords. No keywords were reported from 9 papers by the keyword extraction algorithm extracted, and those papers were therefore excluded from the analysis, giving a validated sample size of 1363 papers. As shown in Figure 2, we can see that UbiComp has gradually grown from a small conference to become a venue of substantial size.

Of the 66 keywords, mobile device has the highest frequency (N = 692), followed by sensing (N = 453), application (N = 445), design (N = 410), systems (N = 403), location based (N = 366) and data analysis (N = 287). As shown in Figure 3, we created a word cloud to visualize the popularity difference of these keywords. In other words, the field of UbiComp research can be described as a community of researchers who are focused on studying mobile device regarding its sensing, application, design and system issues.

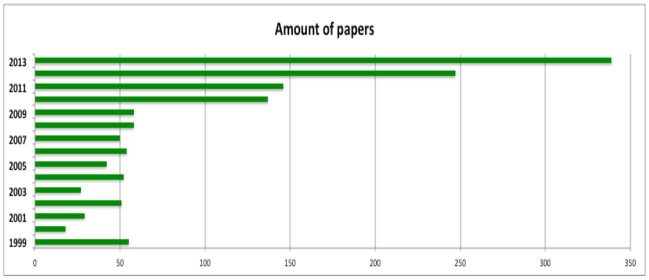


Figure 2. Number of publications per year at UbiComp.

Figure 3. Keyword cloud (1999-2013).

Evolution of the field over time: Correspondence analysis

To visualize how the topics of Ubicomp research evolved over time, we performed a correspondence analysis [23] using a Term-Year binary matrix. This analysis can reveal how different research topics allied with the focus of different years of the conference (Figure 4). More similar keywords are located close to each other, while the distance between publication years and keywords shows how likely certain keywords were published at a specific year. As shown in Figure 4, there is a clear trace showing that consecutive years are located close to each other, and there is no abrupt change between consecutive years. In other words, the field has been mildly but constantly evolving and moving toward new topics.

In addition, as indicated by the diagonal in Figure 4, the years of the conferences form two clusters: the period 1999-2007 and the period 2008-2013. The apparent difference in keyword aggregation in the figure highlights a paradigm shift in research topics in the field. Therefore we separated the sample of papers into these two periods as a basis to investigate the evolution of the field.

During the period of 1999-2007, 378 papers were published in comparison to 985 papers during 2008-2013. In order to focus on the most important keywords and to reduce the noise, our co-word analysis removed about 10 percent of keywords with lowest co-occurrence frequency. As a result, 56 and 59 keywords are retained from the analysis for the two periods respectively.

RESULTS

To identify the research themes in the Ubicomp field, we conducted hierarchical clustering analysis on a correlation matrix with the retained keywords. We used Ward's method with Squared Euclidean Distance as the distance measurement [53]. We adopted a supervised clustering method to reach as many clusters as possible while maintaining content validity and cluster fitness [29,39]. The 56 keywords for 1999-2007 were classified into 12 clusters, labeled as A1-A12 accordingly (Table 2). Each cluster represents a research theme. Similarly, 13 clusters were identified for the period of 2008-2013, labeled as B1-B13 (Table 3). The top-3 most frequent keywords of each cluster are shown in bold, and are used to label each cluster [29,39].

We then generated a weighted co-word graph, whereby each keyword was represented by a node, and keywords that appeared together in *n* papers were linked with an edge of weight *n*. Using this co-word graph we can calculate a number of metrics for each theme of keywords we previously identified. In **Tables 1 and 2** we show for each theme:

- *Keywords*: the set of keywords that constitute this theme;
- *Size*: the number of keywords in the theme;
- *Frequency*: how often, on average, a keyword in this theme appears in our dataset;
- *Co-word frequency*: how often, on average, two keywords in this theme appear on the same paper;
- *Cohesion coefficient*: measures the extent to which when a keyword of this theme appears on a paper then another keyword of this theme also appears on a paper. Indicates the *similarity or dissimilarity* of keywords in a theme. Themes with *higher cohesion coefficient* are more *developed or bridging* research themes [39];
- *Centrality*: the degree of interaction of a theme with other parts of the network based on the use co-occurrence matrix [46]. We calculate a localized version of this metric by setting K-step reach to be 1.
- *Density*: measures the internal cohesion, or the strength, of the links that tie together the cluster of keywords making up the research theme based on the use co-occurrence matrix [8,26]. To minimize the possible bias caused by the different sample sizes of the two periods, when calculating the overall network density, we rely on a binary version of the keyword co-occurrence matrix. This matrix only uses values 1 ("connected") or 0 ("not connected") to characterize every pair of keywords.

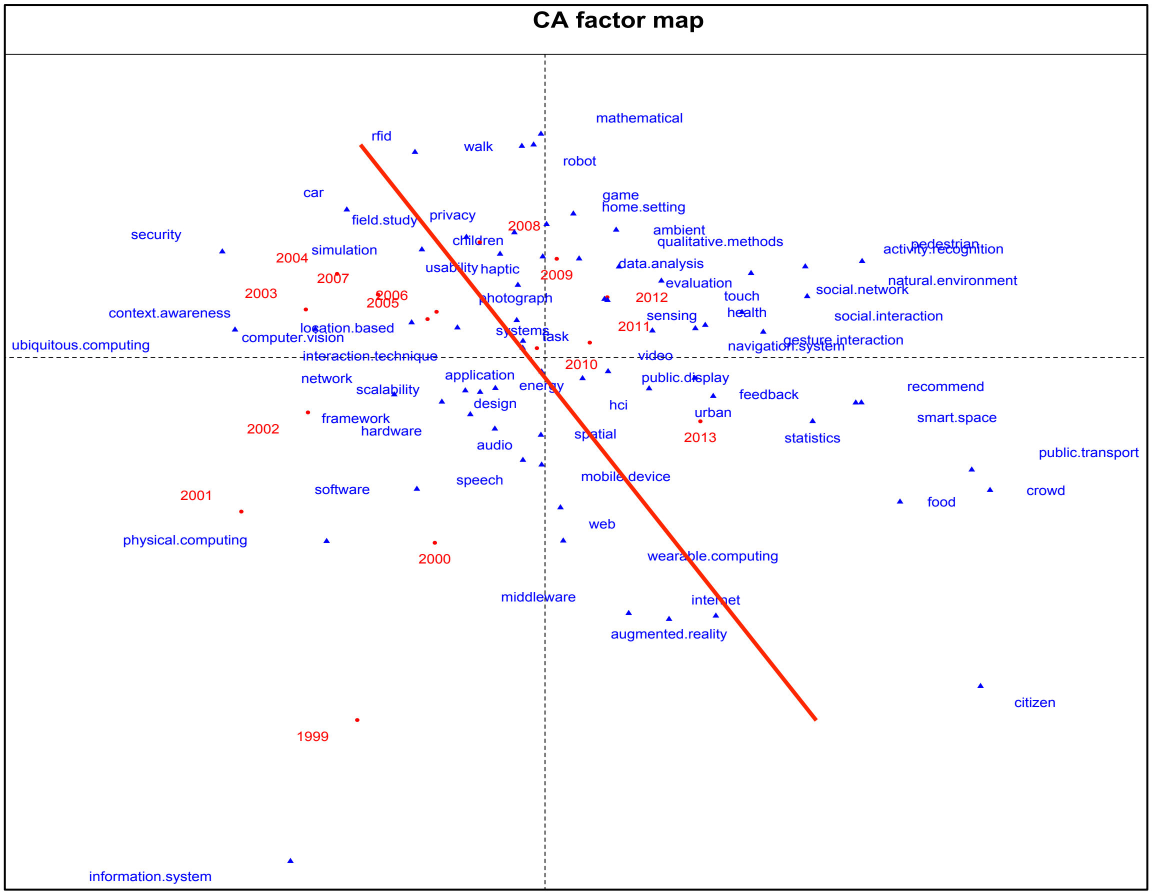


Figure 4. Correspondence analysis of keywords and year. The axes represent the first two principal components.

Based on the values of cluster centrality and density [8,29,39], we constructed two strategic diagrams to visualize the cohesion and maturity of individual research themes, as shown in Figure 5. The origins of these plots are set to the average values of the centrality and density across all the clusters for the designated sample, *i.e.*, (0.869, 5.711) for 1999-2007 and (0.968, 11.133) for 2008-2013. In Figure 6, we show the strategic diagrams of other scientific fields.

To investigate whether the whole research field has become more cohesive in the past 15 years, the overall network density values are calculated for each period. The results show that the overall density of the Ubicomp intellectual map has increased from 0.555 in 1999-2007 to 0.776 in 2008-2013, implying that the research field has become more cohesive while growing.

KEYWORD NETWORK MAPS

For each of the two periods in our dataset we constructed a granular network of keywords using the following procedure. Each keyword is represented as a node in a graph, and we link together keywords that appear together on a paper. In Figures 7 and 8 we show the result of this process for each of the two periods of analysis. We note that in these figures the size of a node is proportional to the frequency of the keyword, and the thickness of links is proportional to the co-occurrence correlation for that pair of keywords. Nodes of the same color belong to the same cluster (Table 1 and 2). To reduce visual clutter we only show a subset of the complete networks, omitting weaker ties and isolated nodes. A downside of this simplification is that, for example, “urban” in Figure 7 appears to be disconnected from its own cluster. This is simply because multiple weaker links are not included.

Next, we focused our analysis on individual keywords rather than underlying themes. A core-periphery analysis was performed to identify the core research topics in the field from the perspective of the whole network structure. Seven keywords (concentration=0.889) and eight keywords (concentration=0.890) were identified to be the core research topics of the whole network in 1999-2007 and 2008-2013 respectively. Keywords or research topics were quantified as follows:

- *Popularity*: how frequently a research keyword is used;
- *Core*: [0-1] how connected is a research keyword with other topics;
- *Structural holes*: how connected is a research keyword with other otherwise distinct topics, thus supporting the topic structure (*i.e.*, the backbone of the field).

A higher core value indicates a keyword that is well connected to other keywords. A higher structural holes count suggests a keyword that brings together otherwise isolated keywords. Keywords with high scores on both of these metrics can be considered as the driving force for advancements in the field: without these keywords, the field of ubiquitous computing would be fragmented. We show these results in Tables 5 and 6.

ID	Keywords	Size	F	CW-F	Cohesion	Centr.	Density
A1	**mobile device, systems, network,** internet	4	102.75	463.50	1.040	1.000	29.67
A2	**sensing, home setting, health,** activity recognition	4	43.50	205.20	1.164	0.942	7.67
A3	**application, context awareness, framework**, ambient, robot, computer vision	6	39.33	179.50	0.608	0.980	3.93
A4	**design, wearable computing, audio**, field study, children	5	37.60	167.40	0.883	0.961	3.90
A5	**location based, data analysis, statistics,** haptic	4	45.25	231.75	1.252	0.904	7.83
A6	**evaluation, task, usability**, speech, interaction technique	5	25.60	129.40	1.584	0.961	3.50
A7	**video, urban, feedback**, game, gesture interaction, car	6	22.33	87.50	1.326	0.960	2.20
A8	**social network, energy, qualitative methods**, software, hardware	5	29.40	122.60	1.002	0.961	3.00
A9	**privacy, security, hci,** social interaction	4	14.50	82.25	0.994	0.808	2.67
A10	**web, walk, navigation system,** public display	4	9.75	49.50	1.327	0.673	1.17
A11	**physical computing, spatial, recommend**, augmented reality, information system	5	11.00	48.80	0.795	0.667	1.00
A12	**simulation, rfid, mathematical**, scalability	4	10.25	58.75	1.299	0.615	2.00

Table 1. Research Themes of 1999-2007: size, frequency (F), co-word frequency (CW-F), cohesion, centrality (Centr.), density.

ID	Keywords	Size	F	CW-F	Cohesion	Centr.	Density
B1	**mobile device, application, systems**, network, energy, web, crowd, scalability	8	173.35	857.50	1.519	1	35.86
B2	**sensing, data analysis, activity recognition**, mathematical	4	165.50	802.50	1.075	1	35.67
B3	**design, social network, home setting**, qualitative methods, privacy, usability, ambient, security, field study, public display	10	80.50	361.80	0.880	1	7.56
B4	**location based, recommend, walk**, rfid	4	80.75	381.25	0.893	1	7.00
B5	**evaluation, task, framework, audio,** speech	5	79.20	419.20	1.682	1	8.70
B6	**video, navigation system, augmented reality**, car, photograph	5	46.20	212.00	1.077	0.981	4.10
B7	**context awareness, smart space, robot**, middleware	4	35.50	170.00	1.080	0.945	3.00
B8	**urban, natural environment, citizen**	3	48.67	235.00	1.191	0.982	5.33
B9	**health, wearable computing, food**	3	54.00	239.00	0.706	0.946	7.67
B10	**feedback, gesture interaction, game,** touch	4	41.25	196.25	1.041	0.927	5.67
B11	**software, hardware**	2	41.00	206.50	1.348	0.912	17.00
B12	**internet, social interaction, physical computing**	3	35.67	165.67	0.979	0.911	4.67
B13	**hci, spatial, simulation,** statistics	4	31.50	146.50	0.881	0.982	2.50

Table 2. Research Themes of 2008-2013: size, frequency (F), co-word frequency (CW-F), cohesion, centrality (Centr.), density.

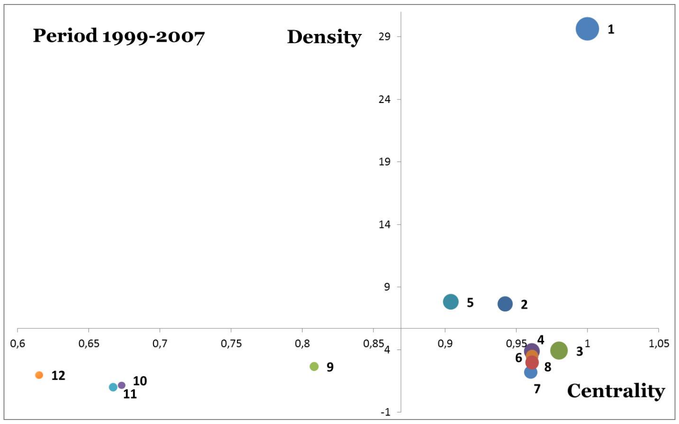


Figure 5a. Strategic diagram of the field in 1999-2007.

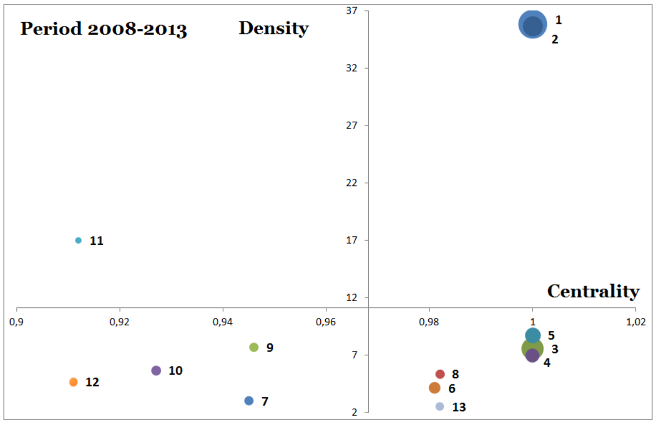
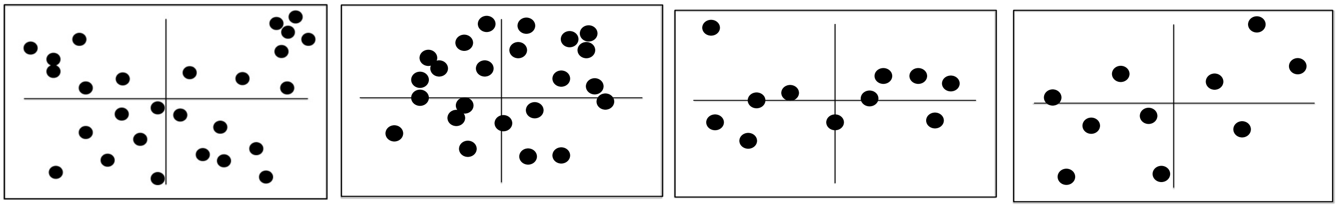


Figure 5b. Strategic diagram of the field in 2008-2013.



Psychology [34] Consumer Behavior [43] Software Engineering [15] Stem Cell Research [3]

Figure 6. Indicative strategic diagrams from other scientific disciplines.

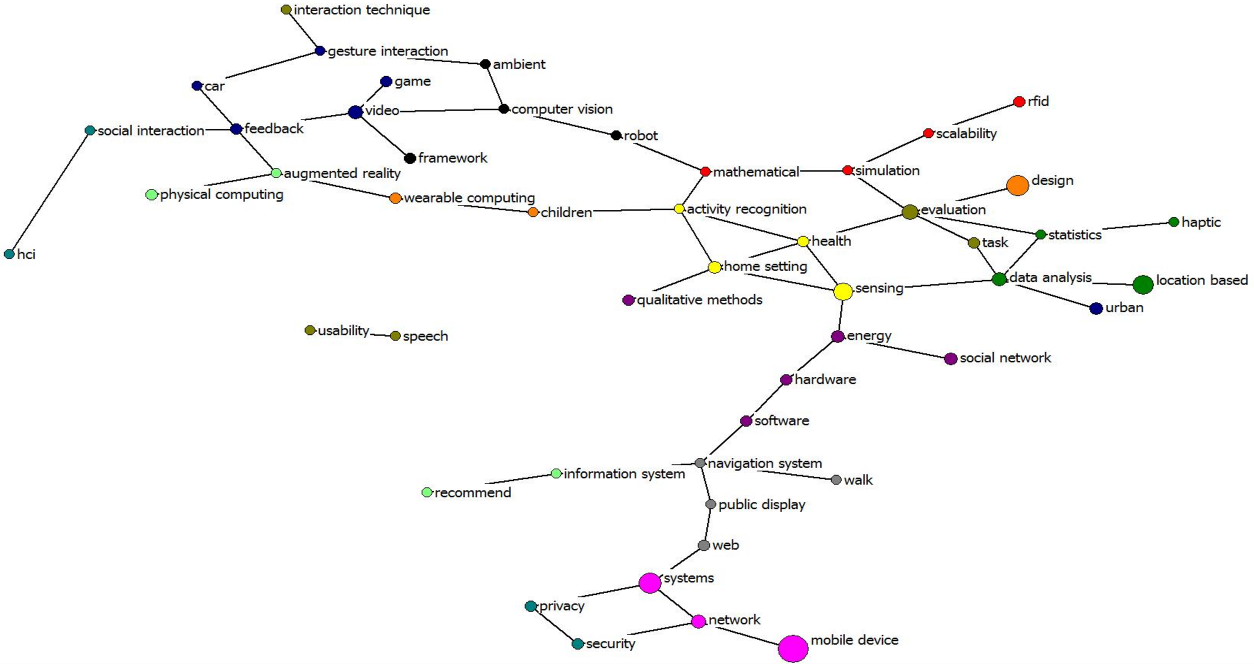


Figure 7. Keywords networking map 1999-2007 (the line represents the link between two keywords with correlation coefficient > 0.13). An interactive version of this graph is available at http://goo.gl/aYknbv.

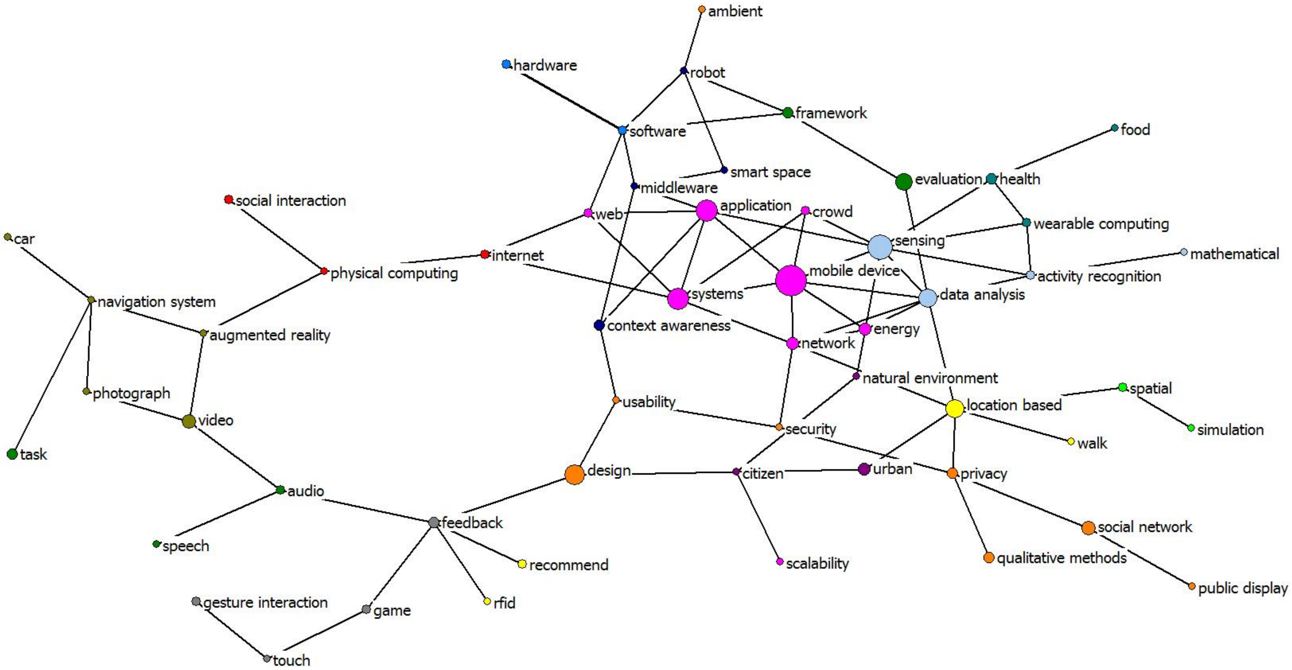


Figure 8. Keywords networking map 2008-2013 (the line represents the link between two keywords with correlation coefficient > 0.09). An interactive version of this graph is available at http://goo.gl/vWmQOA.

No.	Topics	Popularity	Topics	Coreness	Topics	Structural holes
1	mobile device	214	mobile device	0.537	mobile device	37.96
2	application	146	application	0.359	application	34.07
3	design	133	location based	0.333	design	33.86
4	location based	128	design	0.303	location based	32.08
5	systems	115	systems	0.300	evaluation	31.71
6	sensing	106	sensing	0.269	sensing	31.53
7	evaluation	74	network	0.208	systems	31.50

Table 5. Topics with high popularity, coreness, and structural holes during 1999-2007.

No.	Topics	Popularity	Topics	Coreness	Topics	Structural holes
1	mobile device	478	mobile device	0.515	mobile device	39.83
2	sensing	347	sensing	0.375	design	38.34
3	application	299	application	0.329	sensing	38.08
4	systems	288	systems	0.305	systems	37.45
5	design	277	data analysis	0.260	evaluation	36.64
6	location based	238	location based	0.241	location based	36.60
7	data analysis	231	design	0.229	application	36.49
8	evaluation	208	evaluation	0.216	framework	35.75

Table 6. Topics with high popularity, coreness, and structural holes during 2008-2013.

DISCUSSION

The motivation of our paper was to provide empirical evidence on the growth and state of Ubicomp as a research field. Recently, our community has debated the state and future of the field and conference [1], mostly relying on tacit knowledge and subjective (albeit intuitive) insight.

Given the volume of papers published during each period we analysed, the field witnessed an explosive growth of our research community. Due to the constantly low acceptance rate (mostly about 20%), approximately 42 papers were published annually in 1999-2007, and the number increased 4-fold to 164 papers per year in 2008-2013. There is no doubt that UbiComp is now a prospering conference.

In this process, we find that Ubicomp research actually borrows a number of hot topics of other relevant fields, such as social networks and privacy. Therefore, it comes as no surprise that the findings of many Ubicomp papers, which are often multidisciplinarity efforts, are also publishable in other venues, such as Social Psychology and Social Media, and Human-Computer Interaction. This result seems to be consistent with the hypothesis raised by Abowd [1], who argued that Ubicomp is suffering from an identity crisis (at least partly) due to that "*papers [published in the UbiComp conference] could have rightly appeared in one or more other conferences or journals that do not align by name with ubiquitous or pervasive computing.*"

Because of this "lack of uniqueness", one could argue that the field is losing its identity, much like a drop of ink fades in a glass of water. However, our results show that during the past 15 years the whole Ubicomp field has become an increasingly cohesive scientific discipline: we find an overall increase of average cluster centrality (from 0.869 in 1999-2007 to 0.968 in 2008-2013) and density (from 5.711 in 1999-2007 to 11.133 in 2008-2013). This result suggests that although Ubicomp papers are not unique anymore, the field is actually becoming increasingly cohesive: we observe an increasing network density and average cluster centrality and density of the past 15 years. Thus, we argue that rather than thinking of Ubicomp as a drop of ink in a glass of water, we can think of it as a burning candle that shines on and inspires other disciplines. If the Ubicomp field was fading, we would expect a fragmentation of the field and a decrease of connections between topics. However, we observed the contrary.

Underlying trends in ubiquitous computing research

In this section, we attempt to identify evidence that can help us understand how the field has evolved, and at the same time act as a yardstick that can help us compare this field to other fields. We begin our discussion by presenting some of the underlying trends in the Ubicomp field that our analysis has identified.

1999-2007

The strategic diagram in Figure 5 shows that during 1999-2007 the clusters located in quadrant I were: A1 (mobile device, systems, network, internet), A2 (sensing, home setting, health, activity recognition) and A5 (location based, data analysis, statistics) were located in quadrant I. These clusters had high centrality and density values, implying they were driving topics during the crucial first years of our field. In particular, cluster A1 had the highest frequency, centrality and density values apparently outperforming all other clusters. This suggests that the research on systems [48], network [10] and Internet aspects [42] of mobile devices were well-located at the center of the field, driving the advance of other research topics at the same time. In hindsight, we observe that during period 2008-2013, "Internet" separated from cluster A1 to join cluster B12 (Internet, social interaction, physical computing). The relatively big size of A1, A2 and A5 in Figure 5a indicates that they have received substantial attention from the community.

In quadrant III, we observed clusters A9 (privacy, security, hci), A10 (web, walk, navigation system), A11 (physical computing, spatial, recommend) and A12 (simulation, RFID, mathematical). These clusters have low density, low centrality, and low frequency. In other words, the research topics in these clusters received limited attention from researchers during the first years of our field, and were in a state of flux. Indeed, in hindsight, we found that all these clusters have all been disintegrated in 2008-2013, with some of these topics being absorbed by other clusters.

In quadrant IV, clusters A3 (application, context awareness, framework), A4 (design, wearable computing, audio), A6 (evaluation, task, usability), A7 (video, urban, feedback) and A8 (social network, energy, qualitative methods) have high centrality but low density. The high centrality indicates researchers were interested in these topics even if relevant research was still at the initial stage and not yet mature (as suggested by the low density). For example, work on video in cluster A7 was at its infancy [45], and in later years the research on this topic has shifted its focus to navigation system [36] and augmented reality [49]. In hindsight, we found that clusters A3, A4, A7 and A8 have been broken down and reorganized to form new research themes in 2008-2013. Cluster A6 formulated the basis of cluster B5 (evaluation, task, framework).

Interestingly, our analysis did not find any highly developed but marginal research themes existing in the first period (Quadrant II). This was largely expected, given that the field was relatively young and had not developed any well-defined niches during the early years.

2008-2013

In Figure 5b we find two clusters B1 (mobile device, systems, network) and B2 (sensing, data analysis, activity recognition) in quadrant I. The high centrality and density of both clusters indicates that they are the motor themes of the field in recent years. These two clusters are respectively derived from the themes A1 and A2 in 1999-2007, implying the whole field achieved a relative stability regarding driving topics. However, we also observed that internal changes have occurred in these two themes as well, as we will discuss later.

In quadrant II, we observe cluster B11 (software, hardware). This cluster represents the formulation of a highly developed but relatively marginal theme in ubiquitous computing research. In other words, research on software and hardware issue has achieved a relatively high level of maturity, but these research are relatively out of the center of the whole field. In addition, the small total frequency of this cluster implies that they have received limited attentions in the field. This finding is very much in line with a public debate initiated by Landay regarding the favoring of novelty over rigorous systems work [20]. Our findings suggest that perhaps this is indeed the case, with research on systems & hardware becoming a niche topic on the periphery of Ubicomp's knowledge map.

Clusters B7 (context awareness, smart space, robot), B9 (health, wearable computing, food), B10 (feedback, gesture interaction, game) and B12 (internet, social interaction, physical computing) are found in quadrant III. These clusters have both low centrality and low density, and are therefore more likely to change in the near future. Cluster B9 (n=54) has the highest average frequency, suggesting that it has attracted considerable research attention and is more likely to grow than other themes in quadrant III. In B9, we also observed that the keyword food is not really found during the first period of study, indicating an increased sensitivity of our community on issues related to food, health, and personal monitoring, for example monitoring diabetes [44], tracking caloric expenditure [35], and even mobile health games for adults [24].

In quadrant IV, we find clusters B3 (design, social network, home setting), B4 (location based, recommend, walk), B5 (evaluation, task, framework), B6 (video, navigation system, augmented reality), B8 (urban, natural environment, citizen) and B13 (hci, spatial, simulation) with high centrality and low density. In particular, clusters B3, B4 and B5 are well-located at the center of research of the field, considering their high centrality values. The high average frequency of topics in these 3 clusters indicates they have attracted considerable attention, for example an increasing number of papers are beginning to focus on the home setting for design challenges [18], to provide location-based recommendations [31], and developing frameworks that are evaluated in the field [27,55]. Our analysis suggests that these themes have the potential to become future motor themes in the future.

Trending topics

Here we focus our discussion on notable keywords in order to more precisely map their role and evolution over time. We achieve this through a structural hole and core-periphery analysis.

1999-2007

During the period 1999-2007 seven keywords emerged as core topics of the field. In particular, six of these seven keywords are found to be popular, core and backbone topics simultaneously (see Table 5). This suggests a high consistency between research interests, knowledge acquired, as well as effort to maintain the field. The topic "evaluation" is found to be both popular and a backbone topic, but is not a core topic. This suggests that evaluation-focused research was important to maintain the structure of the research field as a whole, but it was not well connected to other topics. This suggests that this topic served to bridge otherwise disparate topics, but it was not well connected.

2008-2013

In this period, eight keywords are at the core topics of the field, seven of which are identified as top popular, core, and backbone topics simultaneously. Similar to the first period, the field exhibits strong consistency of research interests, knowledge acquired, as well as effort to maintain the field as a whole. The topic "framework" appears to be an important backbone and core topic, despite its relatively lower frequency (frequency=63). Therefore, we argue that framework-relevant research is important to bridge a diversity of research topics.

Fading research interest

For each period, our analysis dropped 10 percent of keywords with the lowest co-occurrence frequency. A change in the keywords constituting the 'drop' lists can indicate the tendency of both emerging and fading research topics of the field. For the period of 1999-2007 we dropped 8 keywords that eventually became mainstream during the period 2008-2013: crowd, touch, food, middleware, photograph, smart space, citizen, and natural environment.

For the period of 2008-2013 we dropped five keywords that were mainstream during the period 1999-2007: computer vision, haptic, children, interaction technique and information systems. Our analysis suggests that these topics are fading from the field. Therefore, future research on the recent 'dropped' topics may suffer from a reduced interest from the field.

The topics of pedestrian and public transport are in the 'drop' list for both periods. This implies that despite the constant research effort on these two topics, they are not attractive enough to make them widely popular. Overall, the large shift in the 'drop' lists implies a quick change of research interests in the field at the ground level, albeit while core topics, like mobile devices, sensing and applications, remain consistently at the top.

Fluxionary Research

Research on mobile devices and applications has long been the driving force of the field's advance, and clusters A1 and B1 have a constantly high value of centrality and density in the field. We also observe evolution taking place within this driving theme as well. The topic of application is now included within the cluster B1, while the A1 topic of Internet is showing signs of fading.

A conspicuous growth of themes occurred within cluster B2 (sensing, data analysis, activity recognition), which is a descendant of cluster A2. The fast growth of A2 (and later B2) is a driving force comparable to cluster A1 (the later B1) in 2008-2013. Therefore, cluster B2 to a large extent represents the current momentum of the field's direction.

Location-based research (A5) has substantially changed during the past 15 years. As shown in table 4a, during 1999-2007, location-based research has been a motor theme in the field and relevant studies have been performed through the approach of data analysis on statistics [25]. However, this research direction changed dramatically in recent years. Specifically, the research theme not only gradually drops the topic of haptic, but also shifted from a statistic-driven data analysis approach to be a more service-driven method (recommend, walk, RFID) [54]. Due to these changes, the cluster A5 has been dropped from a motor theme to be a transversal research field with low density.

On the other hand, the topics of data analysis and statistics from A5 merged with the cluster B2 (sensing, data analysis, activity recognition) and B13 (HCI, spatial, simulation, statistics). In other words, data analysis and mathematics-driven studies are becoming increasingly important in the context of sensing research [32].

Cluster B9 (health, wearable computing, food) appears to be an emerging research direction. The same applies to cluster B11 (software, hardware) as well. We notice that cluster B11 has quickly evolved to be a mature research theme during 2008-2013, despite its peripheral position within the field.

Given the lack of big and mature research themes standing out in the field, the decision to merge the UbiComp and Pervasive conferences is likely to speed up the development of the field. We argue that splitting a conference is warranted when there co-exist several highly developed but isolated research themes. Apparently, Ubicomp research is far from reaching such a state.

Limitations

The UbiComp conference has a constantly low acceptance rate, so most papers that were submitted to UbiComp may eventually appear somewhere else. Our sample did not include these papers, which may lead to a bias in sample collection. Also, our keyword extraction algorithm involved substantial supervision from the researchers, which may systematically bias the results. In addition, like other clustering techniques, our co-word analysis can result in clusters with arguably dissimilar keywords. Furthermore, citations of the UbiComp papers may offer important information, which was not included in this paper.

CONCLUSION

Our results reject the assertion that the Ubicomp research field is suffering from an identity crisis. We conclude that the potential for publishing ubiquitous computing research across various venues should not be regarded as evidence of an identity crisis because i) the ubiquitous research community is growing, ii) relevant research has been well focused (especially on mobile devices and sensing), and iii) the whole field has become increasingly cohesive.

During 1999-2007, research interest was mostly on mobile and location-aware applications (*e.g.*, [13,30]), dealing with network challenges of heterogeneity and instability (*e.g.*, [37]) and understanding users' context [16]. Between 2008-2013, although mobile and context are still core topics, research interest in location-aware applications and networks has faded, shifting to sensing (*e.g.*, [17,21,33]), crowdsensing (*e.g.*, [12,22]) and data analysis (*e.g.*, [4,38]).

Given the rapid increase of sensors available in mobile devices, we predict that sensing-related studies will become increasingly core topics of Ubicomp research in the coming few years. As researchers have easier access to data and increasingly reliable network access, researchers are now spending more time on data analysis (*i.e.*, machine learning, prediction) to develop novel technologies.

ACKNOWLEDGMENTS

Funded by the Academy of Finland project 137736, TEKES project 2932/31/2009, and EU grant agreement PCIG11-GA-2012-322138.

REFERENCES

1. Abowd, G.D. What next, ubicomp? Celebrating an intellectual disappearing act. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12.* New York, New York, USA, 2012, 31-40.
2. An, X.Y. and Wu, Q.Q. Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics 88,* 1 (2011), 133-144.
3. An, X.Y. and Wu, Q.Q. Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics 88,* 1 (2011), 133-144.
4. Arase, Y., Ren, F. and Xie, X. User Activity Understanding from Mobile Phone Sensors. In *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing - Adjunct.* ACM, 2010, 391-392.
5. Bauin, S., Michelet, B., Schweighoffer, M.G. and Vermeulin, P. Using bibliometrics in strategic analysis:“understanding chemical reactions” at the CNRS. *Scientometrics 22,* 1 (1991), 113-137.
6. Borgatti, S.P. and Everett, M.G. Models of core/periphery structures. *Social networks 21,* 4 (2000), 375-395.
7. Callon, M., Courtial, J.-P., Turner, W.A. and Bauin, S. From translations to problematic networks: An introduction to co-word analysis. *Social science information 22,* 2 (1983), 191-235.
8. Callon, M., Courtial, J.-P. and Laville, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics 22,* 1 (1991), 155-205.
9. Cambrosio, A., Limoges, C., Courtial, J.P. and Laville, F. Historical scientometrics? Mapping over 70 years of biological safety research with coword analysis. *Scientometrics 27,* 2 (1993), 119-143.
10. Castro, P., Chiu, P., Kremenek, T. and Muntz, R. A Probabilistic Room Location Service for Wireless Networked Environments. In *Ubicomp 2001: Ubiquitous Computing.* Springer Berlin Heidelberg, 2001, 18-34.
11. Chang, P.-L., Wu, C.-C. and Leu, H.-J. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics 82,* 1 (2010), 5-19.
12. Chon, Y., Lane, N.D., Kim, Y., Zhao, F. and Cha, H. Understanding the Coverage and Scalability of Place-centric Crowdsensing. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 2013, 3-12.
13. Couderc, P. and Kermarrec, A.-. Enabling Context-Awareness from Network-Level Location Tracking. In *Handheld and Ubiquitous Computing.* Springer Berlin Heidelberg, 1999, 67-73.
14. Coulter, N., Monarch, I. and Konda, S. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science 49,* 13 (1998), 1206-1223.
15. Coulter, N., Monarch, I. and Konda, S. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science 49,* 13 (1998), 1206-1223.
16. Dey, A.K. Understanding and Using Context. *Personal Ubiquitous Comput 5,* 1 (2001), 4-7.
17. Dey, A.K., Wac, K., Ferreira, D., Tassini, K., et al. Getting closer: an empirical investigation of the proximity of user to their smart phones. In *Ubicomp.* ACM, 2011, 163-172.
18. Dillahunt, T.R. Communication Around Home-energy Monitoring Devices: Connecting Stakeholders in Low-income Communities. In *Proceedings of the 13th International Conference on Ubiquitous Computing.* ACM, 2011, 503-506.
19. Ding, Y., Chowdhury, G.G. and Foo, S. Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management 37,* 6 (2001), 817-842.
20. DUB For the Future: I give up on CHI/UIST. http://dubfuture.blogspot.fi/2009/11/i-give-up-on-chiuist.html, retrieved 24/02/2014.
21. Ferreira, D., Ferreira, E., Goncalves, J., Kostakos, V. and Dey, A.K. Revisiting Human-Battery Interaction with an Interactive Battery Interface. In *Ubicomp.* ACM, 2013, 563-572.
22. Goncalves, J., Ferreira, D., Hosio, S., Liu, Y., et al. Crowdsourcing on the spot: altruistic use of public displays, feasibility, performance, and behaviours. In *Ubicomp.* ACM, 2013, 753-762.
23. Greenacre, M.J. *Correspondence analysis in practice.* Chapman & Hall/CRC, Boca Raton, 2007.
24. Grimes, A., Kantroo, V. and Grinter, R.E. Let's Play!: Mobile Health Games for Adults. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing.* ACM, 2010, 241-250.
25. Guan, Z., Li, Y., Wang, H. and Dai, G. A Comparison of Free Interaction Modes for Mobile Layout System. In *Handheld and Ubiquitous Computing.* Springer Berlin Heidelberg, 2000, 63-71.
26. He, Q. Knowledge Discovery Through Co-Word Analysis. *Library trends 48,* 1 (1999), 133-159.
27. Helaoui, R., Riboni, D. and Stuckenschmidt, H. A Probabilistic Ontological Framework for the Recognition of Multilevel Human Activities. In *Proceedings of the 2013 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing. ACM, 2013, 345-354.

28. Hu, C.-P., Hu, J.-M., Deng, S. and Liu, Y. A co-word analysis of library and information science in China. *Scientometrics Online fir,* (2013), 1-14.
29. Hu, C.-P., Hu, J.-M., Deng, S. and Liu, Y. A co-word analysis of library and information science in China. *Scientometrics Online fir,* (2013), 1-14.
30. José, R. and Davies, N. Scalable and Flexible Location-Based Services for Ubiquitous Information Access. In *Handheld and Ubiquitous Computing.* Springer Berlin Heidelberg, 1999, 52-66.
31. Kalnikaite, V., Rogers, Y., Bird, J., Villar, N., et al. How to Nudge in Situ: Designing Lambent Devices to Deliver Salient Information in Supermarkets. In *Proceedings of the 13th International Conference on Ubiquitous Computing.* ACM, 2011, 11-20.
32. Khan, S.S., Karg, M.E., Hoey, J. and Kulic, D. Towards the Detection of Unusual Temporal Events During Activities Using HMMs. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* ACM, 2012, 1075-1084.
33. Lane, N.D., Lu, H., Eisenman, S.B. and Campbell, A.T. Cooperative Techniques Supporting Sensor-Based People-Centric Inferencing. In *Pervasive Computing.* Springer Berlin Heidelberg, 2008, 75-92.
34. Le Marc, M., Courtial, J.-, Senkovska, E.D., Petard, J.- and Py, Y. The dynamics of research in the psychology of work from 1973 to 1987: From the study of companies to the study of professions. *Scientometrics 21,* 1 (1991), 69-86.
35. Lester, J., Hartung, C., Pina, L., Libby, R., et al. Validated Caloric Expenditure Estimation Using a Single Body-worn Sensor. In *Proceedings of the 11th International Conference on Ubiquitous Computing.* ACM, 2009, 225-234.
36. Li, M., Mahnkopf, L. and Kobbelt, L. The Design of a Segway AR-Tactile Navigation System. In *Pervasive Computing.* Springer Berlin Heidelberg, 2012, 161-178.
37. Lim, H. and Kim, C. Ad-hoc Network Routing for Centralized Information Sharing Systems. In *Handheld and Ubiquitous Computing.* Springer Berlin Heidelberg, 1999, 386-388.
38. Lin, J., Xiang, G., Hong, J.I. and Sadeh, N. Modeling People's Place Naming Preferences in Location Sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing.* ACM, 2010, 75-84.
39. Liu, G.-Y., Hu, J.-M. and Wang, H.-L. A co-word analysis of digital library field in China. *Scientometrics 91,* 1 (2012), 203-217.
40. Liu, X., Zhang, S., Wei, F. and Zhou, M. Recognizing named entities in tweets. In *Human Language Technologies.* Association for Computational Linguistics, 2011, 359-367.
41. Liu, Y., Goncalves, J., Ferreira, D., Xiao, B., et al. CHI 1994-2013: Mapping two decades of intellectual progress through co-word analysis. In *CHI.* ACM, Toronto, Canada, 2014.
42. Molina-Jiménez, C. and Marshall, L. Anonymous and Confidential Communications from an IP Addressless Computer. In *Handheld and Ubiquitous Computing.* Springer Berlin Heidelberg, 1999, 383-385.
43. Muñoz-Leiva, F., Viedma-del-Jesús, M.I., Sánchez-Fernández, J. and López-Herrera, A.G. An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality & Quantity 46,* 4 (2012), 1077-1095.
44. Nachman, L., Baxi, A., Bhattacharya, S., Darera, V., et al. Jog Falls: A Pervasive Healthcare Platform for Diabetes Management. In *Pervasive Computing.* Springer Berlin Heidelberg, 2010, 94-111.
45. Neustaedter, C. and Greenberg, S. The Design of a Context-Aware Home Media Space for Balancing Privacy and Awareness. In *UbiComp 2003: Ubiquitous Computing.* Springer Berlin Heidelberg, 2003, 297-314.
46. Nielsen, A.E. and Thomsen, C. Sustainable development: the role of network communication. *Corporate Social Responsibility and Environmental Management 18,* 1 (2011), 1-10.
47. Porter, M.F. and Street, C.E. An algorithm for suffix stripping. *Program: electronic library and information systems,* (1980).
48. Randell, C. and Muller, H. Low Cost Indoor Positioning System. In *Ubicomp 2001: Ubiquitous Computing.* Springer Berlin Heidelberg, 2001, 42-48.
49. Rashid, Z., Nur, K., Carreras, A. and Pous, R. Browsing Reality: Dynamic Contextualization in Human Scale Smart Spaces. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication.* ACM, 2013, 13-16.
50. Ritzhaupt, A.D., Stewart, M., Smith, P. and Barron, A.E. An Investigation of Distance Education in North American Research Literature Using Co-word Analysis. *International Review of Research in Open & Distance Learning 11,* 1 (2010).
51. Rombach, M.P., Porter, M.A., Fowler, J.H. and Mucha, P.J. Core-periphery structure in networks. *arXiv preprint arXiv:1202.2684,* (2012).
52. Vaughan, L., Yang, R. and Tang, J. Web co-word analysis for business intelligence in the Chinese environment. *Aslib Proceedings 64,* 6 (2012), 653-667.
53. Ward Jr, J.H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association 58,* 301 (1963), 236-244.
54. Yuan, J., Zheng, Y., Zhang, L., Xie, X. and Sun, G. Where to Find My Next Passenger. In *Proceedings of the 13th International Conference on Ubiquitous Computing.* ACM, 2011, 109-118.
55. Zheng, J. and Ni, L.M. An Unsupervised Framework for Sensing Individual and Cluster Behavior Patterns from Human Mobile Data. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* ACM, 2012, 153-162.