Hot Topics in CHI: Trend Maps for Visualising Research

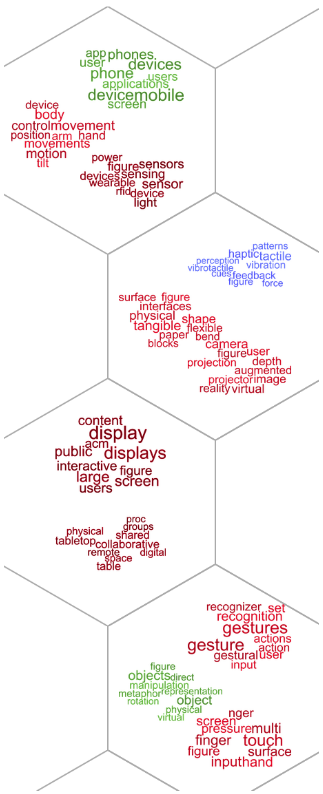

Figure 1. Clipped section of the at a glance view of the HCI area.

Stefano Padilla
Heriot-Watt University
Edinburgh,
Scotland, UK EH14 4AS
S.Padilla@hw.ac.uk

Thomas S. Methven
Heriot-Watt University
Edinburgh,
Scotland, UK EH14 4AS
T.Methven@hw.ac.uk

David W. Corne
Heriot-Watt University
Edinburgh,
Scotland, UK EH14 4AS
D.W.Corne@hw.ac.uk

Mike J. Chantler
Heriot-Watt University
Edinburgh,
Scotland, UK EH14 4AS
M.J.Chantler@hw.ac.uk

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

ACM 978-1-4503-2474-8/14/04…$15.00.
http://dx.doi.org/10.1145/2559206.2578867

Abstract
The aim of this paper is to introduce a novel method of identifying and visualising research trends in an automated, unbiased way. The output of this we call a 'Trend Map', and in this paper we use it to present an at-a-glance overview of the CHI research area, showing which areas are 'hot', 'cold', and 'stable'. This specimen Trend Map was created using the past five years of CHI publications as our only input. We hope that providing this at-a-glance overview of the recent CHI area will encourage introspection and discussion within the community.

Author Keywords
Trend Map; topic modelling; self-organising maps; visualisation; research portfolio; CHI; introspection; automated modelling; discussion; at-a-glance overview

ACM Classification Keywords
H.5.m. [Information interfaces and presentation]: Miscellaneous.

Introduction
Human-computer interaction (HCI) has been growing continually over the last decade. It is a complex, diverse, and interdisciplinary area which covers a very broad spectrum of research [12]. As such, researchers

Figure 2. The CHI conferences included in our Trend Map.

and industrialists new to the area can often find the breadth and scope of the research daunting.

In last year's most prominent HCI conference, CHI 2013 [8, 20], PhD students and industry accounted for 62% of all attendees [4]. We feel that providing an at-a-glance overview of the whole area would be particularly valuable to these types of participants. The timing is also serendipitous as we believe this goal embraces CHI 2014's theme "*one of a CHInd: exploring the challenges and possibilities of engagement and partial connections across different disciplines, communities, and purposes at CHI*" [3].

There have been previous studies into the HCI area including the development of Taxonomies [24], analysis of authors [5, 14], and visual explorations of the area [11]. These studies have either been performed manually (and as such are time consuming to repeat or recreate) or are not focussed on the research interests of CHI directly. Our motivation is, instead, to create an automatically generated overview of the CHI community which elucidates normally hidden information such as important topics, trends, and research affinities, showing which topics in CHI are 'hot' or 'cold'.

Importantly, we are not going to consider CHI sessions as a method of examining trends for two reasons. First, the CHI sessions change year-to-year and make multi-year comparisons difficult. Second, we believe that papers often do not easily fit into a single topic; instead they are composed of and contribute to multiple different fields and interests.

With these goals in mind, we defined three core tenets we wished our visualisation technique to fulfil. The overview should:

- be at-a-glance and fit legibly on a single A4 page;
- be generated automatically with minimal author selected parameters (see *Table 2*); and
- elucidate normally hidden information such as important topics, trends, and research affinities.

To this end, in this paper we present one possible solution of providing an overview of CHI, a visualisation which we call a 'Trend Map'. This Trend Map is a novel way of presenting a synopsis of an entire research area on a single A4 page or standard 1080p monitor. By using a combination of automated topic modelling and dimensionality reduction, we create a hexagonal grid of coloured word clouds. This grid allows the reader to see the most important topics within CHI, their affinity and relationships to each other (with closer topics being more similar to distant ones), and how their popularity has been trending over the last five years (*Figure 2*).

The full CHI Trend Map is presented on the final page in *Figure 10,* with an interactive version available on the web[1]. As the work presented in this paper is designed to engage discussion and introspection within the CHI community, we welcome any feedback or suggestions from the reader.

[1] http://bit.ly/Hot-Topics-In-CHI

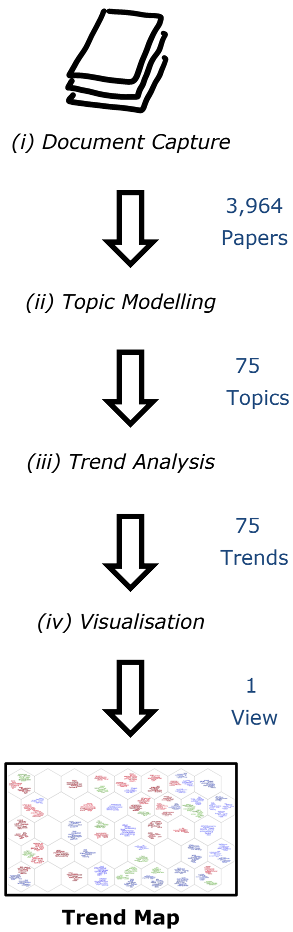

Figure 3. The four necessary steps to create a Trend Map

Creating Trend Maps

Before we discuss the CHI trends presented on the last page, we will describe in detail the four steps used to create it. These are outlined in *Figure 3* and described in detail below:

(i) Document Capture
As Trend Maps are automatically generated, first we defined and prepared our corpus. Our corpus constitutes the proceedings, including the extended abstracts, of the last five years of CHI. The papers and abstracts were downloaded from the ACM Digital Library [2] as PDF documents. In total 3,964 documents were retrieved.

Raw text was extracted automatically from each PDF document and stored as simple, unformatted text files. The measured extraction failure rate for the documents was very small (0.6%) and, on inspection, this was mostly caused by incorrect font embedding. In addition, we removed stop words, capitalisation, numbers and symbols from the corpus.

(ii) Topic Modelling
Once the input data was captured and processed, we extracted the research concepts from the papers. This has been previously been done manually. For example, CHI has been categorised into many sub-disciplines by way of eight Programme Committee subcommittees [1]. In addition, previous studies have tried to cluster CHI papers into categories [27, 21]. One of our goals, however, was to automate the process as much as possible, so it was repeatable, free from predefined labels, and low cost.

Consequently, we used topic modelling [6] as:

- topics can be discovered across large corpuses,
- the number of topics can be predefined,
- it works with unstructured data, and
- it has been proven to work well with scientific corpora [10, 15, 19, 25].

We utilised topic modelling with Latent Dirichlet Allocation as defined by McCallum [18]. We reduced the CHI corpus to 100 different topics (LL/token -9.29414), where each topic is a list of 20 labels defining a single research concept. The full list of topics is available online (http://bit.ly/Hot-Topics-In-CHI).

One side effect of the document capture and topic modelling process is that some of the extracted topics did not contribute meaningful research concepts and so were removed. For example:
5: ing tion de con tions inter pro ment dis tive im…
33: rst signi con speci dif cation ed de classi pro bene…

In addition, topics appearing in more than 20% of the corpus were also removed, as they constituted topics that were too generic. The value of 20% was chosen by the author, aided by examination of topic similarity dendrogram. Generic topics were mostly formed from common papers terms. For example:
9: human systems computer proceedings conference…
38: users user interface figure design results page…

After removing both the meaningless topics and the topics of common terms, we were left with 75 topics representing the research interests of CHI.

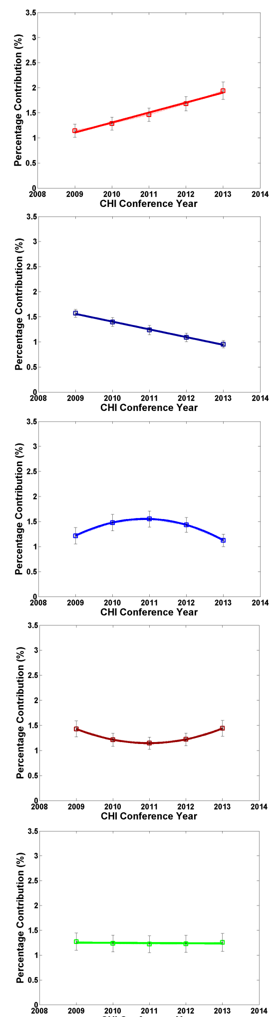

Figure 4. Mean graphs of the five trend categories used

(iii) Trend Analysis

One important aspect of the CHI corpus we wanted to visualise was the trends of topics over the last five years. To do this, we categorised the topics into 'hot', 'cold' and 'stable' groups. The trends we used in our Trend Map were inspired by the various stages of the Gartner Hype Cycle [17]; however, as it is not known if research follows similar stages, we decided to simplify these to just five stages, as shown in *Figure 4*.

Each of our 75 topics was classified into one of the five categories, using the weight information from the topic modelling, polynomial fitting and a simple classification algorithm.

As each of the last five years of CHI had different numbers of papers, we normalised each year's weighting of topics so that every year had equal influence. In addition, the topic contribution for each paper was also normalised, so topic weights in each paper account for one paper even after the common topics were removed.

In topic modelling, each paper can contribute to many topics. For each topic, therefore, we summed the normalised contribution of all the papers for each conference year, resulting in the percentage that topic contributed to CHI that year. We then fitted linear and quadratic polynomials (p^2 > 0.95) to each topic, and calculated the turning point locations of the quadratic fittings.

Finally, the conditions described in *Table 1* were used in an automated classification algorithm to label each topic as one of the five trend types. In our classification, we assumed a rate of change of 0.6% of the normalised corpus per year as the boundary between a topic which has plateaued and one which changing.

Growing (16)		• The turning point (the min or max of the quadratic fit) is outside the five years and the topic is increasing. • The max of the quadratic fit is in the last year and the topic is increasing. • The min of the quadratic fit is in the first year and the topic is increasing.
Sliding (18)		• The turning point is outside the five years and the topic is decreasing. • The max of the quadratic fit is in the first year and the topic is decreasing. • The min of the quadratic fit is in the last year and the topic is decreasing.
Peak (16)		• The turning point is in the core three years and the first term in the quadratic fit is negative.
Trough (13)		• The turning point is in the core three years and the first term in the quadratic fit is positive.
Plateau (12)		• The turning point is outside the five year and the topic is not increasing or decreasing faster than 0.6%. • The turning point is inside the five years and not increasing or decreasing faster than 0.6%.

Table 1. Conditions for the automatic classification of topics

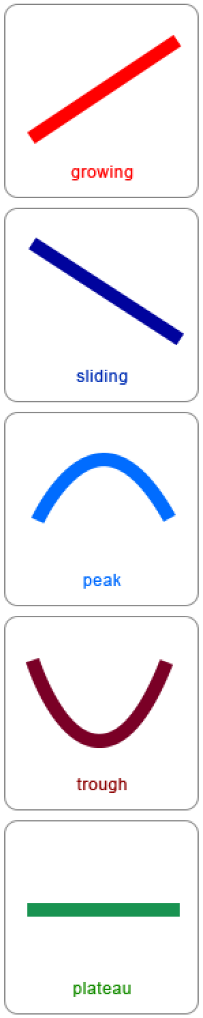

Figure 5. The colours assigned to each type of trend.

(iv) Visualisation

A word cloud for each topic was created using IBM's Word Cloud Generator [13]. Red, blue, and green were selected to show which topics were hot, cold, and stable, respectively. We decided at this stage to pragmatically assign each of the five topic classifications to one of these three colours, based on the most recent movement of the topic. Therefore, the growing and trough topics were coloured red, the sliding and peak topics were coloured blue and the plateau topics were coloured green (see *Figure 5*).

The height of each term matches the weight from the topic modelling. The first ten terms from each topic were rendered in each topic cloud; in addition, terms were placed horizontally to aid reading. *Figure 6* shows examples of the final word clouds.

Figure 6. Word clouds representing topics 56, 34 and 87.

Our aim is to display the Trend Map in a single at-a-glance two-dimensional visualisation, with similar topics adjacent to each other. We define similarity between two topics as the level of commonality between their contributing papers[2]. We used a Self-Organising Map (SOM) [16] to create and organise our layout as SOMs are designed to layout similarity data into predefined, geometric shapes.

[2] Similarity is defined as the cosine distance between two topic vectors in the space represented by the 3,964 papers

SOMs rely on a vector quantisation method consisting of clusters organized on a regular low-dimensional grid. The SOM algorithm is similar to *k*-means [9] and multidimensional scaling (MDS) [7], but in contrast, topological neighbours in a SOM are updated and stretched towards similar samples, while maintaining the distance relationships of the underlying data. This method to display data has been previously used in other browsing tasks [22, 23]. In short, the closer items are to each other, the more similar they are, with those displayed in the same shape being most similar.

We implemented the SOM as defined by Vasanto et al [26] and used a hexagonal lattice as the layout to allow for more adjacency between topics. In addition, clusters of topics were displayed on a single level using the distances between topics to decide orientation.

The optimal grid dimension as calculated by the SOM was an 8x5 grid as seen in *Figure 10*. Finally, we rotated the grid to a landscape layout to fit better on 1080p screens.

The Trend Map was then converted into HTML code for publication to the Internet. In addition, we added features to enable a more detailed exploration of the corpus, including a magnifying glass and the raw trend data and fitting for each topic.

Discussion

To reiterate our objectives, the overview should:

- be at-a-glance and fit legibly on a single A4 page;
- be generated automatically with minimal author selected parameters (see Table 2); and

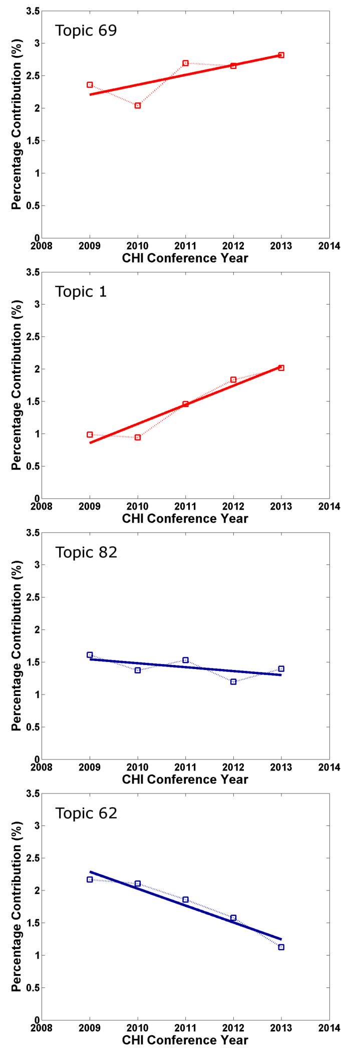

Figure 7. Examples of growing and sliding topics

- elucidate normally hidden information such as important topics, trends, and research affinities.

It is the final point we will focus on in this section. We believe there are two important aspects of this Trend Map (*Figure 10*) that should be discussed. First, we will discuss individual topic trends, and then we will examine the overall layout.

Individual Topic Trends

Before discussing the overall layout of the SOM, we will investigate some of the topic trends which have been highlighted by analysing the previous five years of CHI papers. As discussed previously, there are five different categories a topic could fall into, as shown in *Figure 4*. Due to the quantisation inherent in this process, however, it is inevitable that certain details will be lost. In this section, therefore, we will briefly discuss the different types of trends in both the growing and sliding topics.

The first observation to make is that in the 'growing' topics there is a range of speeds of growth and maturity. For example, topic 69 (touch, finger, multi) and topic 1 (behaviour, mental, technology) are growing at different rates and show different contribution levels, as can be seen in *Figure 7*.

Of course, the same can be said about the 'sliding' areas too, with topic 82 (eye, gaze, tracking) and topic 62 (web, page, pages) decreasing at different rates and from different contribution levels.

At this point, it is worth mentioning that there could be many reasons for the decrease of papers in a certain topic, such as a change in common or fashionable language (such as computer programs now more commonly being known as apps), funding calls causing researchers to change focus (such as the recent Horizon 2020 framework), or even the area becoming big enough to spawn its own special interest group (such as the WWW conference).

The final trends which we will briefly mention are the peaks and troughs. Importantly, as previously discussed, we made a pragmatic decision to class these as hot or cold topics based on the most recent movement of the topic. Of course, this decision can be debated, but we believed this was important to create a good at-a-glance visualisation. There is less variance in these two trends, as the conditions for categorisation are stricter, but they offer interesting insights. For example topic 88 (experience, user, product) was at its lowest contribution in 2010, but has since started rising quickly as shown in *Figure 8*. Conversely, topic 63 (social, Facebook, network) peaked in 2010 and has since starting sliding in contribution.

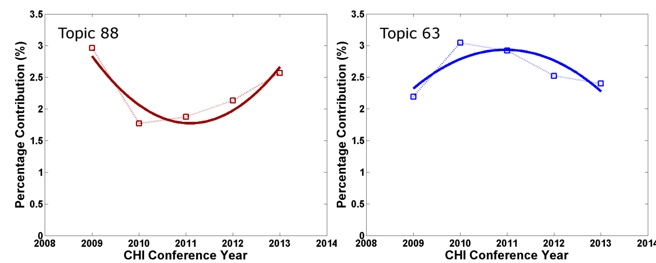

Figure 8. Examples peak and trough topics

This implies that topics which are currently sliding or growing are not guaranteed to continue that way and their contribution could change dramatically in CHI

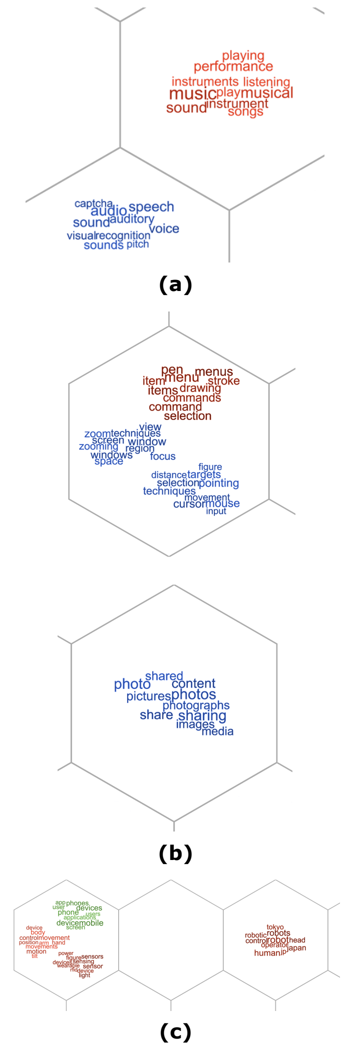

Figure 9. Examples of interesting hexes.

2014. This Trend Map, therefore, should be considered as a snapshot in time, rather than as a prediction of the future. As such, we plan to regularly update the visualisation to include new CHI publications and topics.

Trend Map Layout

For providing an at-a-glance overview of the CHI area, it is the overall layout of the Trend Map which we believe is the most informative. Thanks to the constraints of the SOM, close topics (especially topics in the same hex) will be more similar than distant topics. This leads to both interesting adjacencies and interesting opposites. For example, topic 16 (audio, speech, sound) which is adjacent to topic 52 (music, musical, sound) shows that topics can share words that are used in different contexts, as shown in *Figure 9a*. It seems intuitive that these should be near each other, but it is interesting that the topic modelling has separated these into different topics, implying that perhaps there are two sub areas in CHI which look into similar areas, but from different viewpoints. Another interesting point is that while the audio area appears to be cooling, the music area is heating up. As discussed in the previous section, there could be many reasons for this, but it is an interesting open question and discussion topic whether the authors of the papers in the audio topic have evolved into another area of CHI (such as the music topic) or whether they have given birth to other, more specialised conferences.

To contrast this observation, we can instead look at opposite corners of the SOM to see whether the topics presented are distant and distinct from each other. For example, the bottom left corner of the SOM seems to be concerned with traditional HCI topics such as menus, commands, windows, zooming, pointing, and cursors (topics 0, 36, and 84). In the opposite corner, however, there is a single topic (25) which concerns sharing photos, both of which can be seen in *Figure 9b*. Indeed, it could possibly be argued that the SOM as presented changes relatively smoothly from input methods, design and hardware (the technical underpinnings of CHI) in the bottom left, to what you can do and present to a user with those technologies in the top right. For example, maps and navigation is relatively close to the bottom left, while a topic concerning location, place and activities is close to the top right. The obvious advantage of this, of course, is that it instantly and at-a-glance shows the wide range of both CHI challenges and research even to non-experts in the field. It is also important to remember at this stage that this has emerged automatically from the CHI papers without any external categorisation or input.

Finally, there is the interesting open question of ‘what would fill the gaps?’ As the SOM is a method of laying out items in a 2D space, the fact that gaps appear in it is an interesting discussion point. Any gap is the result of the SOM algorithm not being able to find a topic which would fit in that position. Does this mean there are unexplored areas in CHI? Or perhaps there are possible synergies between topic areas which could be explored? For example, what would fill the space at the top left of the SOM between robots and mobile, sensors, and movement, shown in *Figure 9c*? We leave this as an open point of discussion and would welcome input from the CHI community.

P1	The last **5** CHI conferences as input
P2	Calculated **100** Topics
P3	Removing the top **20%** most common topics
P4	Setting the change boundary at **0.6%**
P5	Using a **hexagonal** SOM
P6	Selected **Red**, **Blue** and **Green** for trends

Table 2. Author selected parameters for discussion

Conclusion

To conclude, we will briefly recap the advantages of using a Trend Map. The Trend Map is:

- an at-a-glance overview of a research area;
- generated automatically with minimal author selected parameters; and
- able to elucidate normally hidden information such as important topics, trends, and research affinities.

As such, we believe that this technique is a valuable method of displaying the complex and often daunting amount of research topics inherent in a research area in a visually appealing way. In addition, we hope that this synopsis of the CHI area encourages introspection and discussion on the current state of HCI research, how it has developed over the past five years, and where it might go in the future.

This technique is designed to be open and transparent with only minimal author selected parameters, as shown in *Table 2*. One of the biggest advantages, therefore, is that this technique can be easily applied to future CHI conferences and other conferences in the HCI and wider area, thus allowing at-a-glance comparisons of the main foci and trends within different communities.

Finally, we are conscious that this is just one way in which the conference trends could be visualised. We created the CHI Trend Map in order to encourage discussion within the community, and so we welcome any feedback or discussion about the methods and final visualisation presented in this paper. There are no doubt many interesting possible trends and patterns exposed by the Trend Map which we haven't covered in this paper and the authors are interested in hearing other researchers' comments and suggestions as to how the visualisation could be improved.

Acknowledgements

We would like to thank our colleagues for their thoughtful discussions and feedback. Research assisted by EPSRC grant EP/I038845/1.

References

[1] ACM CHI2014. Selecting a subcommittee. Last accessed Dec 2013. http://chi2014.acm.org/authors/selecting-a-subcommittee

[2] ACM Digital Library. CHI Conference. Last accessed Dec 2013. http://dl.acm.org/event.cfm?id=RE151

[3] ACM SIGCHI. CHI2014: One of a CHInd Conference. Last accessed Dec 2013. http://chi2014.acm.org

[4] ACM SIGCHI. CHI by numbers. Last accessed Dec 2013. http://chi2013.acm.org/sample-page/chi-by-numbers/

[5] Bartneck, C., & Hu, J. Scientometric analysis of the chi proceedings. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM (2009), 699-708

[6] Blei, D. M. Probabilistic topic models. *Communications of the ACM*, *55*(4), (2012), 77-84.

[7] Cox, T. F., & Cox, M. A. Multidimensional scaling. *CRC Press*, (2010).

[8] Google. Top Publications – Human Computer Interaction. Last accessed December 2013. http://scholar.google.co.uk/citations?view_op=top_venues&hl=en&vq=eng_humancomputerinteraction

[9] Gray, R. M. Vector quantization. *IEEE ASSP Magazine*, 4-29, (1984).

[10] Griffiths, T. L., & Steyvers, M. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, *101*(Suppl 1), (2004), 5228-5235.

[11] Henry, N., Goodell, H., Elmqvist, N., & Fekete, J. D. 20 years of four HCI conferences: A visual exploration. *International Journal of Human-Computer Interaction*, 23(3), (2007), 239-285.

[12] Hewett, T., Baecker, R., Card, S., Carey, T., Gasen., J., Mantei, M., Perlman, G., Strong, G., and Verplank, W. *ACM SIGCHI Curricula for Human-Computer Interaction*. ACM Press (1996).

[13] IBM. Word cloud generator. Last Access Jan 2011. http://www.alphaworks.ibm.com/tech/wordcloud

[14] Kaye, J. J. (2009, April). Some statistical analyses of CHI. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, ACM (2009), 2585-2594.

[15] Khalifa, O., Corne, D. W., Chantler, M., & Halley, F. Multi-objective Topic Modeling. In *Evolutionary Multi-Criterion Optimization*. Springer Berlin Heidelberg, (2013), 51-65.

[16] Kohonen, T. Self-Organizing Maps. Springer, (2000).

[17] Linden, A., & Fenn, J. Understanding Gartner's hype cycles. *Gartner* (2003).

[18] McCallum, A. K. Mallet: A machine learning for language toolkit. (2002).

[19] Methven, T., Padilla, S., Corne, D. W., & Chantler, M. J. Research Strategy Generation: Avoiding Academic 'Animal Farm'. *In Proc. of 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14).* DOI= 10.1145/2556420.2556785, http://doi.acm.org/10.1145/2556420.2556785. ACM, New York, NY, USA. (2014), *25-28*.

[20] Microsoft. Top conferences in human-computer interaction. Last accessed Dec 2013. http://academic.research.microsoft.com/RankList?entitytype=3&topDomainID=2&subDomainID=12

[21] Newman, W. A preliminary analysis of the products of HCI research, using pro forma abstracts. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1994), 278-284.

[22] Padilla, S., Halley, F., & Chantler, M. J. Improving Product Browsing whilst Engaging Users. *Digital Engagement'11*, (2011), 15-17.

[23] Padilla, S., Halley, F., Robb, D. A., & Chantler, M. J. Intuitive Large Image Database Browsing Using Perceptual Similarity Enriched by Crowds. In *Computer Analysis of Images and Patterns*. Springer Berlin Heidelberg, (2013), 169-176.

[24] Quinn, A. J., and Bederson, B. B., Human computation: a survey and taxonomy of a growing field*. In Proc, of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM (2011), 1403-1412.

[25] Talley, E. M., Newman, D., Mimno, D., Herr II, B. W., Wallach, H. M., Burns, G. A. & McCallum, A. Database of NIH grants using machine-learned categories and graphical clustering. *Nature Methods*, *8*(6), (2011), 443-444.

[26] Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. Self-organizing map in Matlab: the SOM toolbox. In *Proceedings of the Matlab DSP conference*, Vol. 99, (1999), 16-17.

[27] Wulff, W., & Mahling, D. E. An assessment of HCI: issues and implications. *ACM SIGCHI Bulletin*, 22(1), (1990), 80-87.

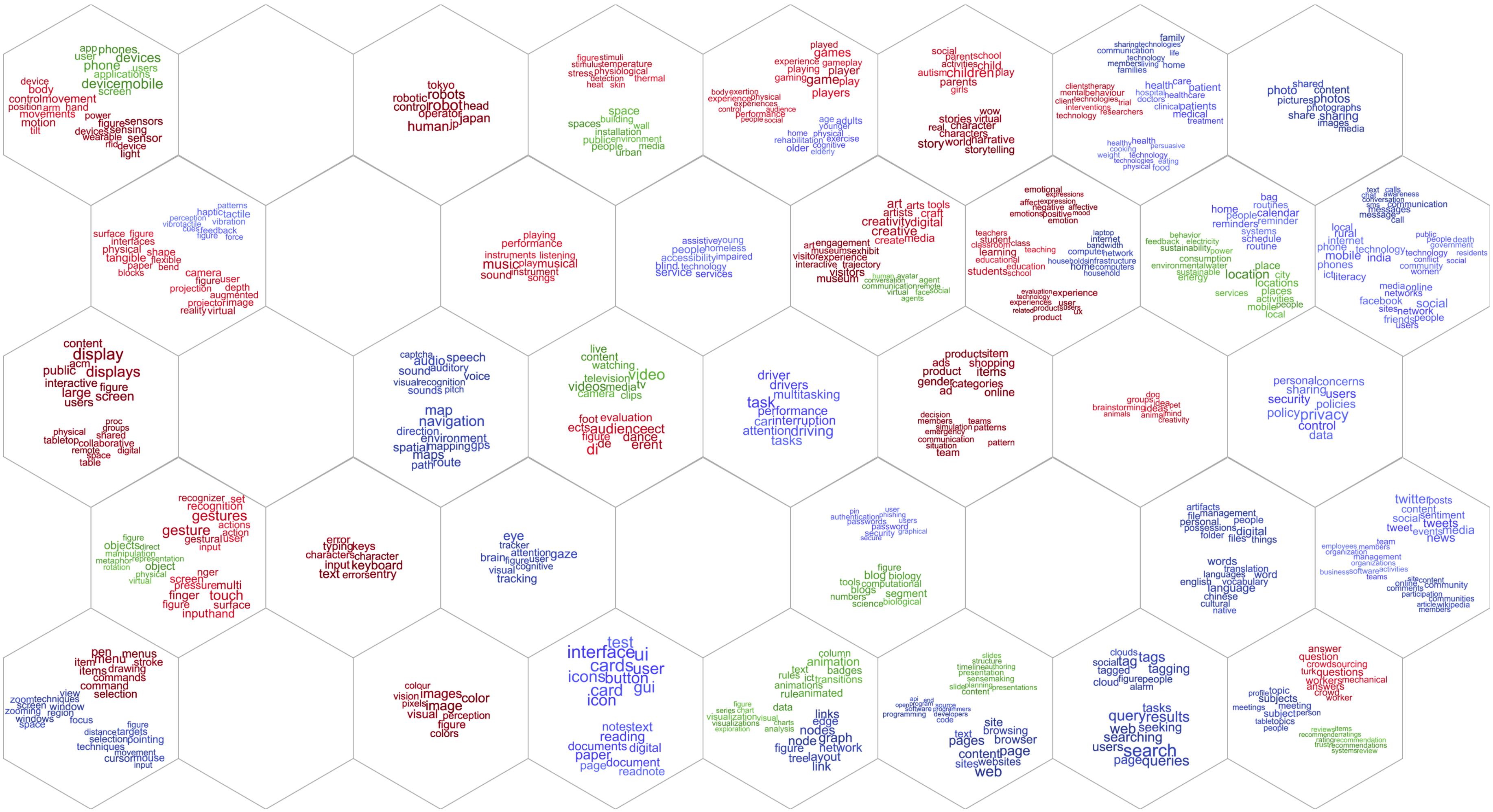

Figure 10. Trend Map showing topics from the last five CHI conferences. Enhanced version and list of topics at http://bit.ly/Hot-Topics-In-CHI