

Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together?

Andreas Holzinger

► **To cite this version:**

Andreas Holzinger. Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together?. Alfredo Cuzzocrea; Christian Kittl; Dimitris E. Simos; Edgar Weippl; Lida Xu. 1st Cross-Domain Conference and Workshop on Availability, Reliability, and Security in Information Systems (CD-ARES), Sep 2013, Regensburg, Germany. Springer, Lecture Notes in Computer Science, LNCS-8127, pp.319-328, 2013, Availability, Reliability, and Security in Information Systems and HCI. <hal-01506781>

HAL Id: hal-01506781

<https://hal.inria.fr/hal-01506781>

Submitted on 12 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together?

Andreas Holzinger^{1,2}

¹ Medical University Graz, Institute for Medical Informatics, Statistics and Documentation
Research Unit Human–Computer Interaction
Auenbruggerplatz 2/V, A-8036 Graz, Austria
a.holzinger@hci4all.at

² Graz University of Technology, Institute for Information Systems and Computer Media
Inffeldgasse 16c, A-8010 Graz, Austria
a.holzinger@tugraz.at

Abstract. A major challenge in our networked world is the increasing amount of data, which require efficient and user-friendly solutions. A timely example is the biomedical domain: the trend towards personalized medicine has resulted in a sheer mass of the generated (-omics) data. In the life sciences domain, most data models are characterized by complexity, which makes manual analysis very time-consuming and frequently practically impossible. Computational methods may help; however, we must acknowledge that the problem-solving knowledge is located in the human mind and – not in machines. A strategic aim to find solutions for data intensive problems could lay in the combination of two areas, which bring ideal pre-conditions: Human–Computer Interaction (HCI) and Knowledge Discovery (KDD). HCI deals with questions of human perception, cognition, intelligence, decision-making and interactive techniques of visualization, so it centers mainly on supervised methods. KDD deals mainly with questions of machine intelligence and data mining, in particular with the development of scalable algorithms for finding previously unknown relationships in data, thus centers on automatic computational methods. A proverb attributed perhaps incorrectly to Albert Einstein illustrates this perfectly: “Computers are incredibly fast, accurate, but stupid. Humans are incredibly slow, inaccurate, but brilliant. Together they may be powerful beyond imagination”. Consequently, a novel approach is to combine HCI & KDD in order to enhance human intelligence by computational intelligence.

Keywords: Human–Computer Interaction (HCI), Knowledge Discovery in Data (KDD), HCI-KDD, E-Science, Interdisciplinary, Intersection science

1 Challenges in the Data Intensive Sciences

Through the continuing exponential growth of data size and complexity, along with increasing computational power and available computing technologies, the data intensive sciences gain increasing importance [1]. E-Science is being advanced as a new science alongside theoretical science, experimental science, and computational science, as a fundamental research paradigm [2]. Meanwhile, it is established as the *fourth paradigm* in the investigation of nature, after theory, empiricism, and computation [3], [4].

One of the grand challenges in our networked 21st century is the large, complex, and often weakly structured [5], [6], or even unstructured data [7]. This increasingly large amount of data requires new, efficient and user-friendly solutions for handling the data. With the growing expectations of end-users, traditional approaches for data interpretation often cannot keep pace with demand, so there is the risk of delivering unsatisfactory results. Consequently, to cope with this rising flood of data, new computational and user-centered approaches are vital.

Let us look, for example, at the life sciences: biomedical data models are characterized by significant complexity [8], [9], making manual analysis by the end users often impossible [10]. At the same time, experts are able to solve complicated problems almost intuitively [11], often enabling medical doctors to make diagnoses with high precision, without being able to describe the exact rules or processes used during their diagnosis, analysis and problem solving [12]. Human thinking is basically a matter of the “plasticity” of the elements of the nervous system, whilst our digital computers (Von-Neuman machines) do not have such “plastic” elements [13] and according to Peter Naur for understanding human thinking we need a different, non-digital approach, one example given by his Synapse-State theory [14].

Interestingly, many powerful computational tools advancing in recent years have been developed by separate communities with different philosophies: Data mining and machine learning researchers tend to believe in the power of their statistical methods to identify relevant patterns – mostly automatic, without human intervention, however, the dangers of modeling artifacts grow when end user comprehension and control are diminished [15], [16], [17], [18]. Additionally, mobile, ubiquitous computing and sensors everywhere, together with low cost storage, will accelerate this avalanche of data [19], and there will be a danger of drowning in data but starving for knowledge, as Herbert Simon pointed it out 40 years ago: “A wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it” [20].

Consequently, it is a grand challenge to work towards enabling effective human control over powerful machine intelligence by the integration of machine learning methods and visual analytics to support human insight and decision support [21], the latter is still *the* core discipline in biomedical informatics [8].

A synergistic combination of methodologies, methods and approaches of two areas offer ideal conditions for addressing these challenges: HCI, with its emphasis on human intelligence, and KDD, dealing with computational intelligence – with the goal of supporting human intelligence with machine intelligence – to discover new, previously unknown insights within the flood of data.

The main contribution of HCI-KDD is, following the notion: “science is to test ideas, engineering is to put these ideas into business” [22], to enable end users to find and recognize previously unknown and potentially useful and usable information. It may be defined as the process of identifying novel, valid and potentially useful data patterns, with the goal of understanding these data patterns for decision support.

2 Human–Computer Interaction in a Nutshell

HCI evolved from computer science interest in input-output technologies [23]; hence it is a relatively young discipline, with its beginnings in the period when computers just began to have input-output devices on perception and cognition in interacting with machines [24]. Because of multimedia applications, the importance of HCI continues to increase [25], [26], [27]. Today, the main focus of HCI research is on human issues, including perception, cognition, human intelligence and interaction with any kind of information at any type of device [28].

A recent and perfect example of HCI research is the area of visual analytics, which has established its own scientific community [29] and maybe had its origins in Shneiderman’s *visual analytics mantra* (“overview first, zoom/ filter, details on demand” [30]) and Keim’s *extended mantra* (“Analyse First – Show the Important – Zoom, Filter and Analyse Further – Details on Demand” [29], [31]).

Keim et al. [29] provided a formal description of the visual analytics process, following some notions of Van Wijk [32], which demonstrates a core research of HCI relevant to solve problems involved with complex data:

Let the input data sets be $S = S_1, \dots, S_m$ whereas each set $S_i, i \in (1, \dots, m)$, consists of attributes A_{i1}, \dots, A_{ik} .

The output of this process will be called **Insight I**. This I is either obtained directly from a set of created **Visualizations V**, or through confirmation of stated **Hypotheses H** as the results of the automatic analysis methods.

Formally this process can be described as a transformation $F: S \rightarrow I$, whereas F is a concatenation of functions $f \in \{D_W, V_X, H_Y, U_Z\}$ defined as follows:

D_W describes the basic pre-processing functionality with $D_W: S \rightarrow S$ and $W \in \{T, C, SL, I\}$, including the data transformation functions D_T , data cleaning functions D_C , data selection functions D_{SL} , and data integration functions D_I , which are required to make the analysis functions applicable to the data set.

$V_W, W \in \{S, H\}$ symbolizes the visualization functions, which are either functions visualizing the data $V_S: S \rightarrow V$ or functions visualizing hypotheses $V_H: H \rightarrow V$.

$H_Y, Y \in \{S, V\}$ represents the hypothesis generating process.

Important in this view is that Keim et al. distinguished between functions that generate hypotheses from data $H_S: S \rightarrow H$ and functions that generate hypotheses from visualizations $H_V: V \rightarrow H$.

Human–Computer Interactions $U_Z, Z \in \{V, H, CV, CH\}$ are thus an integral part in this formal model. Such interactions may either affect visualizations only (i.e. selecting, zooming, ...) $U_V: V \rightarrow V$, or the user interactions can affect hypotheses only by generating new hypotheses on the basis of the given ones $U_H: H \rightarrow H$.

Insight can be gained from visualizations $U_{CV} : V \rightarrow I$ or from the hypothesis itself $U_{CH} : H \rightarrow I$.

This leads to a main research question in HCI: “What is interesting?” [33]. Beale (2007) [34] brings it to point: Interest is an essentially human construct, a perspective on relationships between data influenced by tasks, personal preferences and experience. Interest, like beauty, is in the eye of the beholder and is strongly dependent on previous knowledge and past experience [35]. For a correct semantic interpretation a computer would need to understand the context in which a term is presented; however, a comprehension of complex context is beyond computation. For this and many other reasons we cannot leave the search for knowledge to computers alone. We must be able to guide them as to what it is we are looking for and on which areas to focus their phenomenal computing power. Hence, in order for a data mining system to be generically useful to us, it must have some way in which we can indicate what is interesting, and for that to be dynamic and changeable [36].

Another recent and relevant research route in HCI is Attention Routing, which is a new idea introduced by Chau et al. [37] to overcome one critical problem in visual analytics: to help end users locate good starting points for analysis. Attention is of enormous importance for human information processing and through the process of selective attention, only selected subsets of the vast collection of incoming information becomes designated for further processing [38], [39], [40].

Based on anomaly detection [41], attention routing methods quasi-channel the end-user’s attention through massive data structures, e.g. network structures, towards interesting nodes or sub-graphs that do not conform to “normal” behaviour. Such abnormalities often represent new knowledge that directly leads to novel insights.

From the early beginning, HCI research was focused on the measurement of human performance, heavily borrowing from the cognitive sciences [42],[35]; today HCI-research includes a large collection of various methods [43], [44], [45].

3 Knowledge Discovery in a Nutshell

Maimon & Rokach (2010) [46] define Knowledge Discovery in Databases (KDD) as an automatic, exploratory analysis and modeling of large data repositories and the organized process of identifying valid, novel, useful and understandable patterns from large and complex data sets. Data Mining (DM) is the core of the KDD process [47].

The term KDD actually goes back to the machine learning and Artificial Intelligence (AI) community [48]. Interestingly, the first application in this area was again in medical informatics: The program Rx was the first that analyzed data from about 50,000 Stanford patients and looked for unexpected side-effects of drugs [49]. The term really became popular with the paper by Fayyad et al. (1996) [50], who described the KDD process consisting of 9 subsequent steps:

- 1. Learning from the application domain:** includes understanding relevant previous knowledge, the goals of the application and a certain amount of domain expertise;

- 2. Creating a target dataset:** includes selecting a dataset or focusing on a subset of variables or data samples on which discovery shall be performed;

3. Data cleansing (and preprocessing): includes removing noise or outliers, strategies for handling missing data, etc.);

4. Data reduction and projection: includes finding useful features to represent the data, dimensionality reduction, etc.;

5. Choosing the function of data mining: includes deciding the purpose and principle of the model for mining algorithms (e.g., summarization, classification, regression and clustering);

6. Choosing the data mining algorithm: includes selecting method(s) to be used for searching for patterns in the data, such as deciding which models and parameters may be appropriate (e.g., models for categorical data are different from models on vectors over reals) and matching a particular data mining method with the criteria of the KDD process;

7. Data mining: searching for patterns of interest in a representational form or a set of such representations, including classification rules or trees, regression, clustering, sequence modeling, dependency and line analysis;

8. Interpretation: includes interpreting the discovered patterns and possibly returning to any of the previous steps, as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns and translating the useful ones into terms understandable by users;

9. Using discovered knowledge: includes incorporating this knowledge into the performance of the system, taking actions based on the knowledge or documenting it and reporting it to interested parties, as well as checking for, and resolving, potential conflicts with previously believed knowledge.

The repertoire of methods in KDD is enormous, ranging from classical supervised methods, for example classification trees, pruning, inducers (ID3, C4.5, CART, CHAID, QUEST, and many others), Bayesian networks, regression frameworks (Regression Splines, Smoothing Splines, Locally weighted Regression), Support Vector Machines (Hyperplane Classifiers, Non-Separable SVM Models, etc.), Rule Induction (e.g. LEMI Algorithm), to classical unsupervised methods, including Clustering Algorithms, Distance Measures, Similarity Functions, Evaluation Criteria Measures, etc., Association Rules, Frequent Set Mining, Constraint-based Data Mining (e.g. Apriori, Max-Miner, MultipleJoins, Reorder, Direct, CAP, SPIRIT, etc.), Link Analysis (e.g. Social Network Analysis), to advanced Methods including Evolutionary Algorithms, Reinforcement Learning Methods, Neural Networks, Granular Computing & Rough Sets, Swarm Intelligence, Fuzzy Logic, Graph Entropy, topological Methods etc. etc. [46], [51], [52], [53], [54]. There are sheer endless possibilities of applications to solve various problems in different areas; here only a few recent examples [55], [56], [57], [58], [59], [60]. A tool to test methods, aiming at finding the most appropriate solution for a specific problem can be found here [61].

4 The novelty of combining HCI-KDD

To attack the grand challenges described above, a novel and promising approach is to combine the best of two worlds – HCI-KDD with the main goal of supporting human intelligence with machine intelligence – to discover new, previously unknown insights into data [5].

One could now argue: Why do we need a combination of two fields, each independently large enough? One answer is that the combination enables the provision of mutual benefits for both disciplines and allows problems to be tackled when they cannot be solved with a single disciplinary view. Likewise, cross-disciplinary or inter-disciplinary views fall short: whilst cross-disciplinary approaches [22] are basically researching outside the scope of their component disciplines, without tight cooperation or integration from other relevant disciplines inter-disciplinary approaches are a “mix of disciplines” [62] this is not a clear sentence. Consequently, trans-disciplinary approaches tend towards a beneficial fusion of disciplines, including reaching a business impact: team members from various disciplines work together on a shared problem by the application of shared conceptual frameworks, which draw together concepts, theories and approaches from those disciplines [63], [64], [65].

Such trans-disciplinary research is carried out with the explicit intent to solve multidimensional, complex problems, particularly problems (such as those related to sustainability) that involve an interface of human and natural systems [64].

This brings us back to the proverb [66] mentioned at the end of the abstract:

Computers are incredibly fast, accurate, and stupid.
Human beings are incredibly slow, inaccurate, and brilliant.
Together they are powerful beyond imagination.

A strategic, synergistic and consequent combination of aspects from HCI and KDD exactly addresses this proverb.

5 Conclusion

Let us summarize the essence of the two fields:

Human–Computer Interaction (HCI), deals mainly with aspects of human perception, cognition, intelligence, sense-making and most of all the interaction between human and machine.

Knowledge Discovery from Data (KDD), deals mainly with aspects of machine intelligence, and in particular with the development of algorithms for automatic data mining.

Both disciplines have large portions of deep, unexplored and complementary subfields. A large challenge, and a possible solution to many current problems in the data-intensive sciences, may be found at the intersection of HCI and KDD.

Consequently, a novel approach is to combine HCI & KDD [5] in order to enhance human intelligence by computational intelligence. The main contribution of HCI-KDD is to *enable* end users to *find and recognize* previously unknown and potentially useful and usable information. It may be defined as the process of identifying novel, valid and potentially useful data patterns, with the goal of *understanding these data patterns* [67].

To visualize this, let us look at Figure 1: The domain expert possesses explicit domain knowledge, and by enabling him to interactively look at data sets he may be able to identify, extract and understand useful information in order to gain new, previously unknown knowledge and insight into his data sets [56].

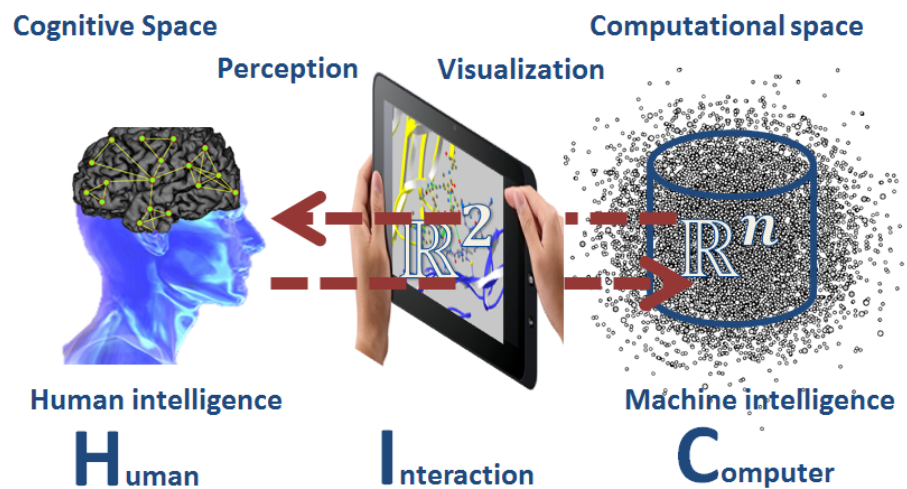


Figure 1. We assume that “thinking” and the formation of hypotheses takes place within the human brain, enabling the interactive analysis of information properties through a synergetic combination of research on human intelligence and computational intelligence [5].

Acknowledgements

I thank the hci4all.at team and all expert members of the task force HCI-KDD.

References

1. Kouzes, R.T., Anderson, G.A., Elbert, S.T., Gorton, I., Gracio, D.K.: The changing paradigm of data-intensive computing. *Computer* 42, 26-34 (2009)
2. Hey, T., Gannon, D., Pinkelman, J.: The Future of Data-Intensive Science. *Computer* 45, 81-82 (2012)
3. Bell, G., Hey, T., Szalay, A.: Beyond the data deluge. *Science* 323, 1297-1298 (2009)

4. Buxton, B., Hayward, V., Pearson, I., Kärkkäinen, L., Greiner, H., Dyson, E., Ito, J., Chung, A., Kelly, K., Schillace, S.: Big data: the next Google. Interview by Duncan Graham-Rowe. *Nature* 455, 8 (2008)
5. Holzinger, A.: On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human-Computer Interaction & Biomedical Informatics. *DATA 2012*, pp. IS9-IS20. INSTICC, Rome, Italy (2012)
6. Holzinger, A.: Weakly Structured Data in Health-Informatics: The Challenge for Human-Computer Interaction. In: Baghaei, N., Baxter, G., Dow, L., Kimani, S. (eds.) *Proceedings of INTERACT 2011 Workshop: Promoting and supporting healthy living by design*, pp. 5-7. IFIP, Lisbon (Portugal) (2011)
7. Holzinger, A., Stocker, C., Ofner, B., Prohaska, G., Brabenetz, A., Hofmann-Wellenhof, R.: Combining HCI, Natural Language Processing, and Knowledge Discovery - Potential of IBM Content Analytics as an assistive technology in the biomedical domain. *Springer Lecture Notes in Computer Science LNCS 7947*, pp. 13-24. Springer, Heidelberg, Berlin, New York (2013)
8. Holzinger, A.: *Biomedical Informatics: Computational Sciences meets Life Sciences*. BoD, Norderstedt (2012)
9. Akil, H., Martone, M.E., Van Essen, D.C.: Challenges and opportunities in mining neuroscience data. *Science* 331, 708-712 (2011)
10. Dugas, M., Schmidt, K.: *Medizinische Informatik und Bioinformatik*. Springer, Berlin, Heidelberg (2003)
11. Polanyi, M.: *Personal Knowledge: Towards a Post-Critical Philosophy*. Nature Publishing Group (1974)
12. Popper, K.R.: *Alles Leben ist Problemlösen*. Piper, München, Zürich (1996)
13. Naur, P.: Computing versus human thinking. *Communications of the ACM* 50, 85-94 (2007)
14. Naur, P.: *The neural embodiment of mental life by the synapse-state theory*. naur. com publishing (2008)
15. Shneiderman, B.: Inventing Discovery Tools: Combining Information Visualization with Data Mining. In: Jantke, K., Shinohara, A. (eds.) *Discovery Science, Lecture Notes in Artificial Intelligence LNAI 2226*, pp. 17-28. Springer Berlin Heidelberg (2001)
16. Shneiderman, B.: Inventing Discovery Tools: Combining Information Visualization with Data Mining. *Information Visualization* 1, 5-12 (2002)
17. Shneiderman, B.: Creativity support tools. *Communications of the ACM* 45, 116-120 (2002)
18. Shneiderman, B.: Creativity support tools: accelerating discovery and innovation. *Communications of the ACM* 50, 20-32 (2007)
19. Butler, D.: 2020 computing: Everything, everywhere. *Nature* 440, 402-405 (2006)
20. Simon, H.A.: *Designing Organizations for an Information-Rich World*. In: Greenberger, M. (ed.) *Computers, Communication, and the Public Interest*, pp. 37-72. The Johns Hopkins Press, Baltimore (MD) (1971)
21. Holzinger, A.: Interacting with Information: Challenges in Human-Computer Interaction and Information Retrieval (HCI-IR). *IADIS Multiconference on Computer Science and Information Systems (MCCSIS), Interfaces and Human-Computer Interaction*, pp. 13-17. IADIS, Rome (2011)
22. Holzinger, A.: *Successful Management of Research and Development*. BoD, Norderstedt (2011)
23. Von Neumann, J.: *The Computer and the Brain*. Yale University Press, New Haven (CT) (1958)
24. Card, S.K., Moran, T.P., Newell, A.: *The psychology of Human-Computer Interaction*. Erlbaum, Hillsdale (NJ) (1983)
25. Helander, M. (ed.): *Handbook of Human-Computer Interaction*. North Holland, Amsterdam (1990)

26. Holzinger, A.: *Multimedia Basics, Volume 2: Learning. Cognitive Basics of Multimedia Information Systems*. Laxmi-Publications, New Delhi (2002)
27. Ebert, A., Gershon, N., Veer, G.: *Human-Computer Interaction. Künstl Intell* 26, 121-126 (2012)
28. Hooper, C.J., Dix, A.: *Web science and human-computer interaction: forming a mutually supportive relationship. interactions* 20, 52-57 (2013)
29. Keim, D., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: *Visual Analytics: Scope and Challenges*. In: Simoff, S., Böhlen, M., Mazeika, A. (eds.) *Visual Data Mining, Lecture Notes in Computer Science* 4404, pp. 76-90. Springer Berlin Heidelberg (2008)
30. Shneiderman, B.: *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336-343 (1996)
31. Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (eds.): *Mastering the Information Age: Solving Problems with Visual Analytics*. Eurographics, Goslar (2010)
32. Van Wijk, J.J.: *The value of visualization*. In: *Visualization, 2005. VIS 05. IEEE*, pp. 79-86. IEEE, (Year)
33. Dervin, B.: *Sense-making theory and practice: an overview of user interests in knowledge seeking and use. J. Knowl. Manag.* 2, 36-46 (1998)
34. Beale, R.: *Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and Web browsing. International Journal of Human-Computer Studies* 65, 421-433 (2007)
35. Holzinger, A., Kickmeier-Rust, M., Albert, D.: *Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society* 11, 279-290 (2008)
36. Ceglar, A., Roddick, J., Calder, P.: *Chapter 4: Guiding Knowledge Discovery through Interactive Data Mining*. In: Pendharkar, P. (ed.) *Managing Data Mining Technologies in Organizations: Techniques and Applications*, pp. 45-86. Idea Group Publishing, Hershey (PA) (2003)
37. Chau, D.H., Myers, B., Faulring, A.: *What to do when search fails: finding information by association. Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 999-1008. ACM, Florence, Italy (2008)
38. Shiffrin, R.M., Gardner, G.T.: *Visual Processing Capacity and Attention Control. Journal of Experimental Psychology* 93, 72-& (1972)
39. Kahneman, D.: *Attention and Effort*. Prentice-Hall, Englewood Cliffs (NY) (1973)
40. Duncan, J.: *Selective attention and the organization of visual information. Journal of Experimental Psychology: General* 113, 501-517 (1984)
41. Chandola, V., Banerjee, A., Kumar, V.: *Anomaly Detection: A Survey. ACM Computing Surveys* 41, (2009)
42. Holzinger, A., Kickmeier-Rust, M.D., Wassertheurer, S., Hessinger, M.: *Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education* 52, 292-301 (2009)
43. Lazar, J., Feng, J.H., Hochheiser, H.: *Research Methods in Human-Computer Interaction*. Wiley, Chichester (UK) (2010)
44. Cairns, P., Cox, A.L. (eds.): *Research Methods for Human-Computer Interaction*. Cambridge University Press, Cambridge (UK) (2008)
45. Nestor, P.G., Schutt, R.K.: *Research Methods in Psychology: Investigating Human Behavior*. SAGE Publications (2011)
46. Maimon, O., Rokach, L. (eds.): *Data Mining and Knowledge Discovery Handbook. Second Edition*. Springer, New York, Dordrecht, Heidelberg, London (2010)
47. Witten, I.H., Frank, E., Hall, M.A.: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco (2011)

48. Piatetsky-Shapiro, G.: Knowledge discovery in databases: 10 years after. *ACM SIGKDD Explorations Newsletter* 1, 59-61 (2000)
49. Blum, R.L., Wiederhold, G.C.: Studying hypotheses on a time-oriented clinical database: an overview of the RX project. *Computer-Assisted Medical Decision Making*, pp. 245-253. Springer (1985)
50. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 39, 27-34 (1996)
51. Piatetsky-Shapiro, G., Frawley, W.: *Knowledge discovery in databases*. MIT Press, Cambridge (MA) (1991)
52. Cios, J., Pedrycz, W., Swiniarski, R.: *Data Mining in Knowledge Discovery*. Academic Publishers (1998)
53. Liu, H., Motoda, H.: *Feature selection for knowledge discovery and data mining*. Springer, Heidelberg, Berlin, New York (1998)
54. Fayyad, U.M., Wierse, A., Grinstein, G.G.: *Information visualization in data mining and knowledge discovery*. Morgan Kaufmann Pub (2002)
55. Billinger, M., Brunner, C., Scherer, R., Holzinger, A., Müller-Putz, G.: Towards a framework based on single trial connectivity for enhancing knowledge discovery in BCI. *Lecture Notes in Computer Science* 7669, pp. 658-667. Springer, Heidelberg, New York (2012)
56. Holzinger, A., Scherer, R., Seeber, M., Wagner, J., Müller-Putz, G.: Computational Sensemaking on Examples of Knowledge Discovery from Neuroscience Data: Towards Enhancing Stroke Rehabilitation. In: Böhm, C., Khuri, S., Lhotská, L., Renda, M. (eds.) *Information Technology in Bio- and Medical Informatics*, vol. 7451, pp. 166-168. Springer, Heidelberg, New York (2012)
57. Holzinger, A., Stocker, C., Peischl, B., Simonic, K.-M.: On Using Entropy for Enhancing Handwriting Preprocessing. *Entropy* 14, 2324-2350 (2012)
58. Holzinger, A., Stocker, C., Bruschi, M., Auinger, A., Silva, H., Fred, A.: On Applying Approximate Entropy to ECG Signals for Knowledge Discovery on the Example of Big Sensor Data. In: R. Huang, e.a.E. (ed.) *Active Media Technologies AMT 2012, LNCS 7669*, pp. 646-657. Springer, Macau (2012)
59. Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Winkler, S., Schaller, S., Holzinger, A.: On text preprocessing for opinion mining outside of laboratory environments. *Active Media Technology, Lecture Notes in Computer Science LNCS 7669*, pp. 618-629. Springer (2012)
60. Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stríteský, V., Holzinger, A.: Opinion Mining on the Web 2.0 – Characteristics of User Generated Content and Their Impacts. *Lecture Notes in Computer Science LNCS 7947*, pp. 35-46. Springer, Heidelberg, Berlin (2013)
61. Holzinger, A., Zupan, M.: KNODWAT: A scientific framework application for testing knowledge discovery methods for the biomedical domain. *BMC Bioinformatics* 14, 191 (2013)
62. Holzinger, A.: *Process Guide for Students for Interdisciplinary Work in Computer Science/Informatics*. Second Edition. BoD, Norderstedt (2010)
63. Moberg, M.: Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. *Futures* 42, 866-873 (2010)
64. Wickson, F., Carew, A.L., Russell, A.W.: Transdisciplinary research: characteristics, quandaries and quality. *Futures* 38, 1046-1059 (2006)
65. Lawrence, R.J., Després, C.: Futures of Transdisciplinarity. *Futures* 36, 397-405 (2004)
66. <http://www.benshoemate.com/2008/11/30/einstein-never-said-that/>
67. Funk, P., Xiong, N.: Case-based reasoning and knowledge discovery in medical applications with time series. *Comput. Intell.* 22, 238-253 (2006)