

# Computer Aided Writing

André Klahold · Madjid Fathi

# Computer Aided Writing

 Springer

André Klahold  
University Siegen  
Siegen, Nordrhein-Westfalen, Germany

Madjid Fathi  
University of Siegen  
Siegen, Nordrhein-Westfalen, Germany

ISBN 978-3-030-27438-2      ISBN 978-3-030-27439-9 (eBook)  
<https://doi.org/10.1007/978-3-030-27439-9>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Contents

<b>1</b>	<b>Introduction</b> .....	1
1.1	Structure of the Book .....	2
<b>2</b>	<b>Writing and Writing Tools</b> .....	5
2.1	Media History .....	5
2.1.1	Origin of Writing .....	5
2.2	Writing Tools .....	8
2.2.1	Handwriting .....	8
2.2.2	The Typewriter .....	8
2.2.3	The First Word Processor .....	10
2.3	Reproduction .....	12
2.3.1	Writing Rooms .....	12
2.3.2	Casting .....	13
2.3.3	Block Printing .....	13
2.3.4	Mobile Bronze Type .....	13
2.3.5	Gutenberg .....	13
2.4	Evolution of Tools .....	14
<b>3</b>	<b>A Model of the Writing Process</b> .....	17
3.1	Task Environment .....	19
3.2	Long-Term Memory .....	19
3.2.1	Planing .....	19
3.2.2	Translating .....	19
3.2.3	Reviewing .....	20
<b>4</b>	<b>Word Processing as Writing Support</b> .....	21
4.1	Capture .....	23
4.1.1	Input .....	23
4.1.2	Navigation .....	24
4.2	Wrapping .....	25
4.3	Correction .....	26

4.4	Layout . . . . .	27
4.5	Automation . . . . .	28
4.6	Semantics . . . . .	29
<b>5</b>	<b>Journalism as a Professional Form of Writing . . . . .</b>	<b>31</b>
5.1	Development Patterns . . . . .	31
5.2	Change of the Medium Through Technology . . . . .	34
5.3	The Beginnings of Computer Aided Journalism . . . . .	34
5.4	Change Through Technology . . . . .	36
<b>6</b>	<b>Research as an Elementary Component . . . . .</b>	<b>37</b>
6.1	Libraries and Efficient Access to Stored Information . . . . .	38
6.2	The Concept of Search . . . . .	38
6.3	Computer-Assisted Searching—Status Quo . . . . .	39
6.4	Knowledge in Multidimensional Representation . . . . .	40
<b>7</b>	<b>Content Management . . . . .</b>	<b>43</b>
7.1	Components of a Content Management System . . . . .	43
7.1.1	Headless CMS . . . . .	43
7.1.2	Core-CMS . . . . .	45
7.1.3	Output-Layer . . . . .	45
7.1.4	Data Storage . . . . .	45
7.2	Media-Neutral Content . . . . .	46
7.3	Structured Content Entry and Management . . . . .	47
7.3.1	An Example . . . . .	47
7.4	Importance of the Data Model . . . . .	49
7.4.1	Runtime Complexity . . . . .	49
7.4.2	Universal Data Model . . . . .	50
7.4.3	Alignment of Structural Changes in the Data Structures Across Systems . . . . .	50
7.4.4	Content, Typography, Structure, Layout . . . . .	51
7.4.5	XML and Typographie . . . . .	51
7.4.6	TML . . . . .	52
7.5	Multimedia Content . . . . .	55
7.6	Multimedia and the Writing Process . . . . .	55
7.6.1	Story Telling . . . . .	58
7.7	Media-Neutral Images . . . . .	58
7.7.1	Media-Neutral Image and Video Editing . . . . .	61
7.7.2	Media-Neutral Semantics on Image Data . . . . .	63
7.7.3	Format Change by Image Adjustment . . . . .	64
7.8	Metadata and the Search for Multimedia Content . . . . .	66
7.9	Media-Neutral Graphics . . . . .	67

- 7.10 Matching of Media-Specific Content Variants . . . . . 67
  - 7.10.1 An Example . . . . . 68
  - 7.10.2 History of “Synchronizing Parallel Versions” . . . . . 69
  - 7.10.3 Simultaneous Editing of Text by Multiple Users . . . . . 70
  - 7.10.4 Pessimistic and Semi-optimistic and Optimistic Locking . . . . . 70
  - 7.10.5 The “Polymorphic Content”-Method (PCM). . . . . 73
  - 7.10.6 Conflict Resolution in the Context of Semi-automatic Locking . . . . . 74
  - 7.10.7 Comparison of Polymorphic Content for Different Media . . . . . 76
  - 7.10.8 Pushing Changes by Automatic Rule-Based Safeguarded Cherry-Picking . . . . . 80
- 8 Knowledge Discovery from Text (KDT) . . . . . 83**
  - 8.1 Keyword Identification . . . . . 83
    - 8.1.1 Refined Keyword Identification for “Brand” Recognition . . . . . 85
  - 8.2 Word Frequency on the Timeline. . . . . 85
  - 8.3 Semantic Relations Between Texts. . . . . 85
    - 8.3.1 Grouping of Texts with Similar Content . . . . . 87
    - 8.3.2 Identifying Topics in Texts . . . . . 88
  - 8.4 Word Associations . . . . . 88
    - 8.4.1 Syntagmatic and Paradigmatic Relationships . . . . . 91
    - 8.4.2 Types of Word Associations . . . . . 93
    - 8.4.3 Quality of Machine-Generated Word Associations . . . . . 94
    - 8.4.4 Coocurents and Collocations . . . . . 95
    - 8.4.5 Asymmetry . . . . . 95
    - 8.4.6 Temporal Aspects. . . . . 97
  - 8.5 Named Entity Recognition (NER). . . . . 97
  - 8.6 Latent Semantic Analysis (LSA) . . . . . 99
    - 8.6.1 Disadvantages of LSA . . . . . 99
    - 8.6.2 A Simple Example of an LSA Application . . . . . 101
    - 8.6.3 Singular Value Decomposition . . . . . 102
    - 8.6.4 Reduction of the Matrix Size . . . . . 106
  - 8.7 Text Summaries . . . . . 107
    - 8.7.1 Adaptive Text Summarization (ATS). . . . . 108
    - 8.7.2 Context-Free Substitution . . . . . 109
    - 8.7.3 The ATS Algorithm . . . . . 109
    - 8.7.4 Context Sensitive Substitution . . . . . 112
    - 8.7.5 Determinitive Composites . . . . . 113
  - 8.8 Multi-topic Detection . . . . . 114
  - 8.9 Ontology Generation . . . . . 114

<b>9</b>	<b>Recommender Systems</b> . . . . .	117
9.1	Recommender Systems and CAW . . . . .	118
9.2	Comparability of the Quality of Recommender Systems . . . . .	119
9.2.1	An Illustrative Example . . . . .	119
9.2.2	Precision, Recall and F-Measure . . . . .	119
9.2.3	Recommendations from Human Experts and the Recall Problem. . . . .	124
9.3	Classification of Recommendation Systems . . . . .	124
9.3.1	Collaborative Filtering . . . . .	125
9.3.2	Content Based Filtering . . . . .	127
9.3.3	Hybrid Methods . . . . .	128
9.4	Distance and Similarity Measures . . . . .	128
9.5	Personalised Recommendation of Unstructured Texts . . . . .	129
<b>10</b>	<b>Automatic Text Generation</b> . . . . .	131
10.1	Template-Based Text Generation . . . . .	131
10.2	The Origins of Automatic Text Generation: T9—Text on 9 Keys . . . . .	132
10.3	Bayesian Networks . . . . .	134
10.3.1	Structure of a Bayesian Network . . . . .	135
10.3.2	Dependencies in the Bayesian Network—a Usecase . . . . .	136
10.4	Markov Chains . . . . .	139
10.5	Hidden Markov Model . . . . .	142
10.6	Recurrent Neural Networks and Long Short-Term Memory Units . . . . .	144
10.6.1	Artificial Neural Networks . . . . .	144
10.6.2	Modern Artificial Neural Networks . . . . .	145
10.7	Learning Artificial Neural Networks. . . . .	147
10.7.1	Backpropagation—Supervised Learning . . . . .	148
10.8	Recurrent Neural Networks (RNN) . . . . .	149
10.9	Long Short-Term Memory RNN (LSTM-RNN) . . . . .	152
10.9.1	Input-Gate . . . . .	153
10.9.2	Forget-Gate . . . . .	153
10.9.3	Output-Gate . . . . .	153
10.9.4	Peephole-LSTM-RNN . . . . .	154
10.9.5	The “Doom”-Boom . . . . .	154
<b>11</b>	<b>Future Developments and Forecasts</b> . . . . .	155
	<b>Bibliography</b> . . . . .	159
	<b>Index</b> . . . . .	167

# List of Figures

Fig. 1.1	Writing model according to Hayes and Flower [1] . . . . .	2
Fig. 1.2	The structure of the book . . . . .	3
Fig. 2.1	Writing tools influence the writing process . . . . .	6
Fig. 2.2	Vinca script on one of the Tartaria tablets . . . . .	7
Fig. 2.3	Examples of Egyptian hieroglyphics . . . . .	7
Fig. 2.4	Evolution from pictorial to syllable writing using the example of hieroglyphics for “thirsty” . . . . .	7
Fig. 2.5	Egyptian hieroglyphs for “water” (around 2000 BC; above) and the letter “Mem” in the Semitic alphabet (around 1800 BC) form the basis of today’s “M”. <i>Source</i> [8] . . . . .	8
Fig. 2.6	Friden Flexo Writer with paper tape feed on the left side. . . . .	9
Fig. 2.7	The paper tape was provided with an 8-bit code. For adjustment purposes there was also a “Feed Hole” in each “line” [15]. . . . .	10
Fig. 2.8	The concept of word processing (and spreadsheets) according to Steinhilper of 1955 [18] . . . . .	11
Fig. 2.9	Copy of the “Jikji” printed in 1377 with movable metal letters. <i>Source</i> French National Library. . . . .	14
Fig. 2.10	Replica of the Gutenberg press at the international printing museum, Carson, California, USA . . . . .	15
Fig. 2.11	The development from cave painting to content management systems [18] . . . . .	16
Fig. 2.12	The dimensions “Writing speed” and “Reproduction”. . . . .	16
Fig. 3.1	The writing model according to Hayes and Flower [1]. . . . .	17
Fig. 3.2	The three phases of the writing process according to Hayes and Flower [1] do not form a strict processing sequence, but influence each other mutually . . . . .	18
Fig. 4.1	Word processing as support for writing . . . . .	21
Fig. 4.2	Overview of the areas in which modern word processing supports the user in the writing process . . . . .	22
Fig. 4.3	“Electronic Pencil” on a Solace-Emulator . . . . .	23



Fig. 4.4	Searching in “Microsoft Word”™ . . . . .	24
Fig. 5.1	Decreasing circulation figures for daily newspapers according to IVW . . . . .	32
Fig. 5.2	A typical theatre with stage and hall. . . . .	32
Fig. 5.3	Film was conserved theater in the early days. The camera took the position of the spectator in the auditorium . . . . .	33
Fig. 5.4	In the cinema, the screen replaced the stage . . . . .	33
Fig. 5.5	The most successful films of all time are at the same time those that have distanced themselves furthest from the theatre through the use of technology. <i>Source</i> <a href="http://www.boxofficemojo.com">http://www.boxofficemojo.com</a> . . . . .	34
Fig. 5.6	Walter Cronkite was one of the pioneers of Computer Aided Journalism at CBS News . . . . .	35
Fig. 6.1	Research as support for the long-term memory of the author . . . . .	37
Fig. 6.2	Relative share of search types in time units in 100% search time . . . . .	39
Fig. 6.3	Basic principle of the MKR framework . . . . .	40
Fig. 6.4	Visualization and filtering in the dimension “Topics” . . . . .	41
Fig. 7.1	The three components of a classic content management system. . . . .	44
Fig. 7.2	Structure of a “headless” CMS . . . . .	45
Fig. 7.3	Output Layer of a “headless” CMS. . . . .	46
Fig. 7.4	Complete graphs with one to five nodes . . . . .	47
Fig. 7.5	Content management as support for structuring in the planning phase . . . . .	48
Fig. 7.6	The output of contents in a media channel takes place in connection with layout, structure and typography . . . . .	51
Fig. 7.7	Only macro typography can be defined via structure . . . . .	52
Fig. 7.8	The TML schematic . . . . .	53
Fig. 7.9	Typography transformation between layouts/media . . . . .	54
Fig. 7.10	Typography transformation via TML . . . . .	54
Fig. 7.11	Multimedia content simplifies the “translation process” during the formulation . . . . .	56
Fig. 7.12	Model for the analysis of text and image comprehension according to Schnotz and Bannert [53]. . . . .	57
Fig. 7.13	Multimedia information processing in simple form (text/image) [52] . . . . .	59
Fig. 7.14	Interactive elements convey complex relationships . . . . .	60
Fig. 7.15	RGB is an additive color model (all colors mixed result in “white”) and is based on the mixture of the colors “Red”, “Green” and “Blue” . . . . .	61

Fig. 7.16	CMYK is a subtractive color model (all colors mixed result in “Black”) of the colors “Cyan”, “Magenta” and “Yellow” and the “Key” as black part. . . . .	62
Fig. 7.17	The RGB colour space almost includes the CMYK colour space . . . . .	62
Fig. 7.18	A focal point of an image and exemplary automated cuts. . . . .	65
Fig. 7.19	Potentially semantically changing image adjustment [61]. . . . .	66
Fig. 7.20	Support for the adjustment of different channel-specific variants . . . . .	67
Fig. 7.21	Adjustment of different channel-specific variants . . . . .	68
Fig. 7.22	Pessimistic locking . . . . .	71
Fig. 7.23	Semi-optimistic locking. . . . .	72
Fig. 7.24	The different forms of locking according to Greenberg [74]. . . . .	72
Fig. 7.25	Unsuccessful conflict management through merge [74]. . . . .	73
Fig. 7.26	The synchronization mechanism in PCM . . . . .	74
Fig. 7.27	Potential for conflict through parallel processing in semi-optimistic versioning. . . . .	75
Fig. 7.28	The version edited by user 1 “v(änd1)” . . . . .	76
Fig. 7.29	The version edited by user 2 “v(änd2)” . . . . .	76
Fig. 7.30	The delta view of changes in comparison. . . . .	76
Fig. 7.31	Different instances of content for different media . . . . .	77
Fig. 7.32	The content as in the original version, from which the instances for medium A and medium B were derived. . . . .	78
Fig. 7.33	The instance “v(änd1)” resulting from modification of “v(org)” . . . . .	78
Fig. 7.34	The instance “v(ext)” resulting from the modification of “v(org)” . . . . .	78
Fig. 7.35	Direct comparison between “v(änd1)” and “v(ext)” . . . . .	78
Fig. 7.36	Edit again from “v(änd1)” to “v(änd2)” . . . . .	78
Fig. 7.37	Differences from “v(änd2)” to “v(ext)” . . . . .	79
Fig. 7.38	Highlighting the changes made in “v(änd2)” . . . . .	79
Fig. 8.1	Keyword analysis based on “Knowledge Discovery from Text” (KDT) . . . . .	84
Fig. 8.2	Knowledge generation on the basis of “Knowledge Discovery from Text” (KDT) . . . . .	86
Fig. 8.3	Recognition of important words in a period of time. . . . .	86
Fig. 8.4	Examples of words of the day: attack in Paris, election in Greece, NASA probe reaches Pluto . . . . .	87
Fig. 8.5	The display of semantically related texts supports the user in the development of knowledge on the topic . . . . .	88
Fig. 8.6	Word clouds depict multi-thematic topics and have a disambiguating effect . . . . .	89

Fig. 8.7	Automatically recognized semantically related texts help with journalistic research. . . . .	89
Fig. 8.8	Grouping of content-related texts in a search result . . . . .	90
Fig. 8.9	Support in finding knowledge on the topic through long-term and short-term associations . . . . .	91
Fig. 8.10	Syntagmatic relations in the “Erlkönig” . . . . .	92
Fig. 8.11	Paradigmatic relations in the “Erlkönig” . . . . .	93
Fig. 8.12	Schematic representation of syntagmatic and paradigmatic relations (from Ulrich [90]). . . . .	94
Fig. 8.13	Temporal memory aspects [94]. . . . .	97
Fig. 8.14	Long-term and short-term associations in a CIMAWA-based implementation . . . . .	98
Fig. 8.15	Support in finding knowledge on the topic through NER. . . . .	98
Fig. 8.16	“Named entities” matching a planned or existing text are automatically determined and displayed . . . . .	100
Fig. 8.17	The semantic intersection of the DBpedia description (left) and the text (right). . . . .	101
Fig. 8.18	Example-Matrix “X”. . . . .	101
Fig. 8.19	Support in “Rewriting” through automatic abbreviation . . . . .	107
Fig. 8.20	Support for correcting fixed text lengths with automated text shortening . . . . .	109
Fig. 8.21	Under- and overfilling of a fixed text box. . . . .	110
Fig. 8.22	The ATS algorithm. . . . .	111
Fig. 8.23	Creation of context-independent synonyms through repeated translation . . . . .	112
Fig. 9.1	Structure of a recommender system . . . . .	117
Fig. 9.2	Support of the writing process through recommendation systems . . . . .	118
Fig. 9.3	The four cases of recommendations . . . . .	120
Fig. 9.4	The four cases of recommendations; the circle visualizes the set of recommended elements . . . . .	121
Fig. 9.5	PRECISION is the percentage of relevant elements (green circular area) in the recommendation (complete circular area). . . . .	122
Fig. 9.6	RECALL is the percentage of relevant elements (complete rectangle) recommended (green circle) . . . . .	123
Fig. 9.7	Search engines present only the most important hits to the user. . . . .	124
Fig. 9.8	Example matrix M on the basis of which collaborative filtering works. . . . .	125
Fig. 9.9	Compressed matrix M with exemplary user profile B <sub>5</sub> . . . . .	126
Fig. 9.10	Matrix M with user profile B <sub>5</sub> and identical users B <sub>12</sub> and B <sub>18</sub> . . . . .	126

Fig. 9.11 Exemplary matrix  $M$  on the basis of which content based filtering works. . . . . 127

Fig. 9.12 The most important distance and similarity measures with interpretation. . . . . 128

Fig. 9.13 In addition to the above visualized recommendation of “suitable” texts for the target group based on their profile, the procedure can also be used to provide the author with the affinity of the text to user profiles already during the writing process . . . . . 129

Fig. 10.1 Automatic text generation in the writing model . . . . . 132

Fig. 10.2 An example of data as basis for text generation (*Source* Narrative Science) . . . . . 133

Fig. 10.3 Push-button telephone “Nokia 3210” with T9 procedure . . . . . 134

Fig. 10.4 D-separable nodes. . . . . 136

Fig. 10.5 Dependencies of word occurrences in a Bayesian network . . . . . 137

Fig. 10.6 Tables of conditional probabilities . . . . . 137

Fig. 10.7 Dependencies with knowledge of the first word . . . . . 137

Fig. 10.8 Changed probabilities when first word is known . . . . . 138

Fig. 10.9 Changed probabilities with knowledge of the last word . . . . . 138

Fig. 10.10 Converging and diverging dependencies . . . . . 139

Fig. 10.11 Simple example of a Markov chain . . . . . 140

Fig. 10.12 Transition matrix of a Markov chain. . . . . 141

Fig. 10.13 A text created with a fifth order Markov chain. . . . . 141

Fig. 10.14 On the basis of the observed state (framed in red) the “real” state can be concluded with certain probabilities . . . . . 142

Fig. 10.15 Principle of a Hidden Markov Model . . . . . 143

Fig. 10.16 Principle of a neural network “McCulloch and Pitts” . . . . . 145

Fig. 10.17 Graph of the logistic function, a special case of the sigmoid function . . . . . 146

Fig. 10.18 Principle of an artificial neural network with sigmoid function. . . . . 147

Fig. 10.19 Gradient descent in the two-dimensional error space . . . . . 149

Fig. 10.20 Principle of a recurrent neural network (RNN) where “KNN” stands for a simple artificial neural network . . . . . 150

Fig. 10.21 Visualization of an RNN as a sequence where “KNN” stand for a simple artificial neural network . . . . . 151

Fig. 10.22 An unfolded RNN where KNN stands for a simple artificial neural network . . . . . 151

Fig. 10.23 An unfold LSTM-RNN where KNN stands for a simple artificial neural network . . . . . 152

Fig. 10.24 An example of a text created with an LSTM . . . . . 154