

UNIVERSITÄT MANNHEIM

THIS THESIS'S TITLE

SUBTITLE

Term Paper

submitted: January 2004

by: Martin Mustermann
born March 28th 1984
in Eberbach

Student ID Number: 012345

University of Mannheim
Chair of Business Administration and Information Systems
D – 68131 Mannheim
Phone: +49 621-181-1691, Fax +49 621-181-1692
Internet: <http://heinzl.bwl.uni-mannheim.de>

Contents

List of Figures	iv
List of Tables	v
List of Abbreviations	vi
1. How to use the template	1
1.1. General structure	1
1.2. Compilation	1
1.3. Languages	2
1.4. Abbreviations	3
1.5. Citations	3
1.5.1. The bib file	3
1.5.2. Citations	4
1.6. Table of Contents	4
1.7. Troubleshooting	4
2. Ein Kapitel	6
2.1. Bilder	6
2.2. Editor	6
2.3. Mathematik	8
3. Summary, Conclusions, and Further Work	13
3.1. Summary	13
3.2. Conclusions	14
3.3. Further Work	15
Bibliography	16

Appendix 17

A. First class of appendices 18

A.1. Some appendix 18

List of Figures

2.1. Das hier steht im Abbildungsverzeichnis.	7
2.2. Dichten univariater Normalverteilungen	10

List of Tables

2.1. Eine sehr einfache Tabelle 10

2.2. Performance of GAs using different types of representations for
deceptive tree problems of different sizes and with different T_{opt}
(arbitrary tree, MST, and star) 11

List of Abbreviations

OWL	Web Ontology Language
TC	telecommunications company

1. How to use the template

This is an evolved latex template. Chapters 2 and 3 remain from former versions of this template. In this first chapter a very short overview on basic handling of this template is given. All users of kile are lucky as configuration instructions are provided only for this latex editor.

1.1. General structure

The structure provided with this template and recommended for use is as follows: The main TeX file “arbeit.tex” is placed in the root directory of the project. All configuration tasks should be done in this main file. This way, one may concentrate on writing rather than bothering with formatting the layout in the single chapters. Figures are placed in the folder ”grafiken“ and the chapters as single files in ”kapitel“. In the directory ”literatur“ there are bibliography styles and the bib file ”lit.bib“ that contains your references. More detailed information on some packages used by the template can be found in directory ”doc“.

1.2. Compilation

This template should be used with latex/pdflatex (text, figures), bibtex (bibliography) and makeindex (table of contents). It is recommendable to use a compilation sequence for the root file as follows: latex (2x), bibtex, makeindex, latex(2x). If you want to create a pdf file just run pdflatex after a full compilation sequence. However, this requires that the figures you use are present in two formats, one that can be handled by latex and one for latex. For example, you can use the eps and the pdf formats.

Furthermore, you must configure your invocation of makeindex. It should look like: ”makeindex arbeit.nlo -s nomencl.ist -o arbeit.nls”. For kile you may configure the options of makeindex in Settings – Configure Kile – Tools – Build –

MakeIndex to `"%S.nlo -s nomenclist -o %S.nls"`. The files `arbeit.nlo` and `arbeit.nls` must be present but may be empty. So if there are no files with the main file's name and the extensions `nlo` and `nls` just create empty ones.

It is convenient to work with dvi files as working copies and to run pdf latex only once in a while to see a pdf file with hyperlinks. DVIs are ugly and have more bugs but they are faster to compile and to render. Moreover, the longer you have stared on an ugly dvi file, the happier you are when the pdf appears and looks nice.

To define a compilation sequence in kile go to Settings – Configure Kile – Tools – Build – QuickBuild. There you may set the sequence, e.g. to: latex, latex, bibtex, makeindex, latex, latex, xdvi (where xdvi is a dvi viewing program).

In short, there are two possible compilation sequences.

Compilation sequence for dvi:

1. latex (2x)
2. bibtex
3. makeindex (make sure it is configured correctly or you will catch odd errors)
4. latex (2x)

Compilation sequence for pdf:

1. latex (2x)
2. bibtex
3. makeindex (make sure it is configured correctly or you will catch odd errors)
4. latex (2x)
5. pdflatex

1.3. Languages

This template is configurable for English and German papers. The entire configuration should be done in the root file `"arbeit.tex"`. To set a language, there are three commands to edit in the root file:

1. the bibliography style at the end of the document must fit the language. For each language there is at least one predefined version you may choose. Look for the "bibliographystyle" tag. Comments are in place explaining what they are about.
2. change the headline of the list of abbreviations to a suitable version. Look for the "nomname" tag.
3. use an appropriate selectlanguage command and remove the other ones.

1.4. Abbreviations

The package nomencl is used to create a list of abbreviations. To add an entry to the list use "abbrev". This is an example for "telecommunications company". To create the list run the compilation sequences as defined above.

1.5. Citations

The citation style used is called natbib and can be used to create numerical and author-year citations. To create the bibliography the bibliography style natdin is used as a basis for German papers, apalike2 for English ones. The respective .sty-file for natdin is located in the directory literatur. More information can be found on <http://merkel.zoneo.net/Latex/natbib.php> or in the directory "doc". Adjust the bib style in the file arbeit.tex to fit your needs. Read the comments, some may help you.

1.5.1. The bib file

The bibliography file is placed in the directory literatur and is called lit.bib. In the lit.bib file you will find different annotations to be used for different kinds of papers. This effects their representation in the bibliography. Some comments can be found in this file, too.

1.5.2. Citations

Talking about a paper one may want to do a citation in the text using `citet` as in Tuttlies et al. (2007). Alternatively parenthesis may be useful (Tannenbaum & Steen, 2007). It is also possible to add more information like the pages of a book one refers to (Tannenbaum & Steen, 2007, pp. 222-333). Padovitz et al. (2008), Mustermann (2009), Welscher (2001a) and Welscher (2001b) are in text citations. There is the Web Ontology Language (McGuinness & van Harmelen, 2005) as an example of an electronic resource.

There are three slightly different bib styles in the directory `literatur`. First, the native `natdin.bst` for DIN citations in German. Second, a slightly modified version `"natdinCustomized"`. This style has minor changes in punctuation and is therefore not conform to DIN 1505. Third, an English translation of the latter is provided. Choose your bibliography style at the end of the root file (`arbeit.tex`). For English papers, `apalike2` looks quite nice.

Several options for citations and the bibliography may be configured in the root file as `usepackage` parameters for `natbib`, e.g. the type of brackets used. See `natnotes.pdf` in the directory `"doc"` for more information. Additionally, for very detailed configuration it is possible to use the file `"natbib.cfg"` in the root directory to override predefined parameters.

1.6. Table of Contents

To create a nice table of contents the package `tocstyle` is used. Configuration options can be found in `tocstyle.pdf` located in `"doc"`.

1.7. Troubleshooting

Use the provided information in log messages. If changes you made do not appear in the document, recompile twice (including `2x latex`, `bibtex`, `makeindex`). Additionally, you may delete all automatically created files, especially the `.aux`, `.bbl`, `.out`, `.toc`, `.lof`, `.lot`, and `.ilg` files. Make sure you do not delete `.tex`, `.nlo`, `.nls`, `.sty`, and `.cfg` files. Moreover, ensure you have used the right compilation

sequence and configured makeindex correctly. If you are using dvi, have a look at a pdf if the problem is also present there, often they are not.

2. Ein Kapitel

Hier kommt eine kurze Zusammenfassung des Kapitels hin.

2.1. Bilder

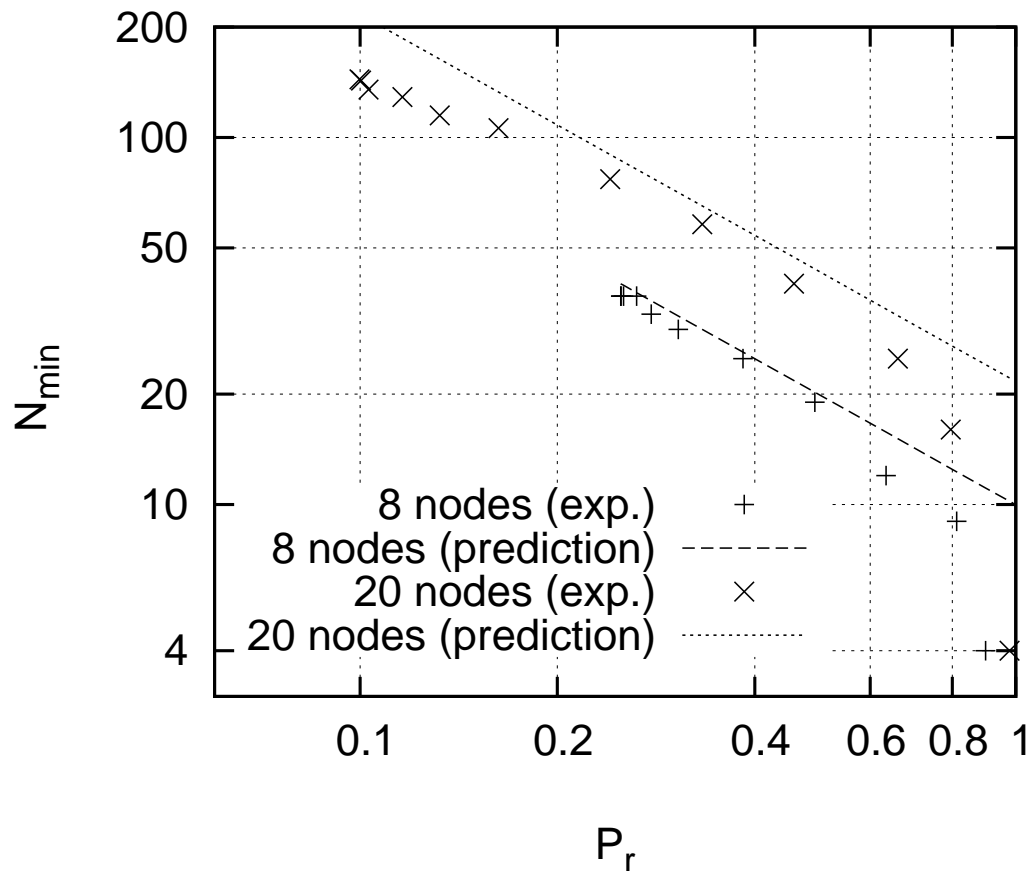
In Abbildung 2.1 findet man ein schönes Beispiel für das Paket subfigure. Subfigure ist sehr praktisch beim Setzen von mehreren Bildern in ein Bild. Man beachte, dass man mit dem Befehl `ref` auf Bilder verweisen kann, wenn diesen mit dem Befehl `label` eine Marke zugewiesen wurde. Nicht alle Dateiformate können so problemlos wie “.eps” mit Latex direkt eingebunden werden. Benötigt man zum Beispiel eine pdf Grafik, so kann man das Dokument mit `pdflatex` compilieren. Hat man alle Bilder sowohl als pdf als auch als eps, kann man wahlweise `latex` oder `pdflatex` nutzen. Ansonsten hilft auch zeitweiliges auskommentieren der Bilder.

Zum Zeichnen von plots ist `gnuplot` sehr zu empfehlen (vgl. Abbildung 2.1(a)). Ist perfekt für zwei oder dreidimensionale Plots und schaut um Welten besser aus als Excel.

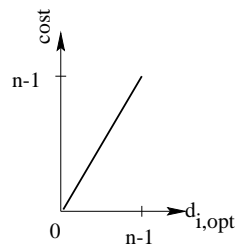
Zum Malen von Bildern ist das Programm `xfig` oder `jfig` sehr gut geeignet (vgl. Abbildung 2.1(b)). Schaut zwar am Anfang ein bißchen seltsam aus, ist aber sehr mächtig.

2.2. Editor

Als Editor für Fortgeschrittene ist `xemacs` gut geeignet. Bei Verwendung von `xemacs` ist der Gebrauch der `reftex` und `auctex`-pakete zu empfehlen. Als Rechtschreibprogramm sollte `aspell` verwendet werden und folgender Code in `init.el` eingefügt werden.



(a) Ein mit gnuplot erzeugtes Bild



(b) Ein mit xfig erzeugtes Bild

Abbildung 2.1.: Beispiele für schöne Bilder.

```

(require 'iso-cvt)
(add-hook 'LaTeX-mode-hook
(function (lambda ()
;; Setze Anführungszeichen etc. fuer Style german
(TeX-run-style-hooks "german")
;;
;; Lade Buffer und wandle nach ISO Latin-1:
(format-encode-buffer 'plain)
)))

(setq ispell-silently-savep t) ;save new words in pdict without questioning
(setq ispell-help-in-bufferp 'electric) ;get a better help buffer
(setq ispell-program-name "aspell")
(setq ispell-extra-args '(W2))

(column-number-mode t)

(custom-set-variables
'(paren-mode 'sexp nil (paren)))

(setq reftex-plug-into-AUCTeX t)
(require 'reftex "reftex" t)
(turn-on-reftex) ; use reftex

```

Neulinge können auch einen sonstigen Editor oder Sachen wie Texnic Center, etc verwenden.

2.3. Mathematik

In latex kann man Formeln wunderschön setzen.

As the following stuff is in English, we must change the hyphenation style.

The OCST problem is defined as follows. Let $G = (V, E)$ be a connected, undi-

rected graph with $n = |V|$ nodes and $m = |E|$ edges. There are communication, or transportation demands, between every pair of nodes. An $n \times n$ *demand matrix* $R = (r_{ij})$ specifies the demands, where r_{ij} is the amount of traffic required between location v_i and v_j . Similarly, an $n \times n$ *distance matrix* $W = w_{ij}$ determines the distance weights between each pair of sites. A tree $T = (V, F)$ where $F \subseteq E$ and $|F| = |V| - 1$ is called a *spanning tree* of G if it connects all the nodes. The weight $w(T)$ of the spanning tree is the weighted sum over all pairs of vertices of the cost of the path between the pair in T . The communication cost over the tree T is defined as

$$w(T) = \sum_{i,j \in V} w_{ij} b_{ij}, \quad (2.1)$$

where $B = b_{ij}$ denotes the traffic flowing directly and indirectly across the edge connecting nodes i and j . It is calculated according to the structure of T . T is the optimal communication spanning tree if $w(T) \leq w(T')$ for all other spanning trees T' . The OCST problem becomes the minimum spanning tree (MST) problem if $w(T) = \sum_{i,j \in V} w_{ij}$. Then, T is the simple minimum spanning tree if $w(T) \leq w(T')$ for all other spanning trees T' .

Cayley's formula identifies the number of spanning trees on n nodes as n^{n-2} . Furthermore, there are n different stars on a graph of n nodes. The similarity between two spanning trees T_i and T_j can be measured using the distance $d_{ij} \in \{0, 1, \dots, n-1\}$ as

$$d_{ij} = \frac{1}{2} \sum_{u,v \in V, u < v} |l_{uv}^i - l_{uv}^j|,$$

where l_{uv}^i is 1 if an edge from u to v exists in T_i and 0 if it does not exist in T_i . The number of edges that two trees T_i and T_j have in common is $n - 1 - d_{ij}$.

Ab jetzt wieder deutsche Trennung.

In Abbildung 2.2 gibt es noch ein Beispiel für ein Bild und in Tabelle 2.1 gibt es ein einfaches Bild. Tabelle 2.2 stellt eine etwas kompliziertere Tabelle dar.

Beim Aufbau von Kommunikationsnetzen stehen Unternehmen vor dem Problem, eine Menge von unterschiedlichen Standorten so durch Leitungen zu verbinden, dass bestimmte Qualitätsanforderungen (z.B. Sicherheit, Bandbreite, Zuverlässigkeit, Skalierbarkeit, etc.) eingehalten werden und gleichzeitig die Netzstruktur so gewählt wird, dass die Kosten des Netzaufbaus minimiert werden. Unternehmen, für welche der Aufbau von Kommunikationsnetzwerken relevant ist, lassen sich

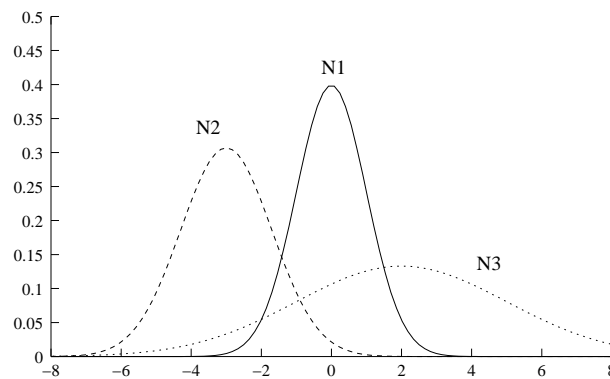


Abbildung 2.2.: Dichten univariater Normalverteilungen

Tabelle 2.1.: Eine sehr einfache Tabelle

linksbündig	zentriert	rechtsbündig
1	2	3,141

in zwei Gruppen unterteilen:

Auf der einen Seite sind Unternehmen zu finden, für die Kommunikationsnetzwerke ein wichtiger Teil ihrer Unternehmensinfrastruktur sind (z.B. Banken, Verlagshäuser, Bahn, Bundeswehr, Unternehmen mit mehreren Niederlassungen, etc.) und welche daher die Kommunikationsnetzwerke oft selbst betreiben. Ein Beispiel hierfür ist die DATEV eG, welche über ein eigenes „Genossenschaftsnetz“ mehrere Dutzend Niederlassungen an die Zentrale in Nürnberg anbindet. Über dieses Kommunikationsnetz wird sowohl der gesamte Daten- als auch Telefonverkehr innerhalb des Unternehmens abgewickelt.

Auf der anderen Seite stehen Unternehmen, für die der Aufbau und der Betrieb von Kommunikationsnetzwerken Kernaufgaben darstellen (z.B. Telekommunikationsunternehmen (Deutsche Telekom oder Deutsches Forschungsnetz (DFN)), Mobilfunkbetreiber (vodafone, T-mobil) oder lokale Provider (NetCologne, Stadtwerke). Für derartige Unternehmen gehört der Aufbau und Betrieb von Kommunikationsnetzwerken zum Kerngeschäft. Netzwerke sind so aufzubauen und weiterzuentwickeln, dass Qualitätskriterien eingehalten werden und gleichzeitig die Betriebskosten minimiert werden.

Beim Aufbau von Kommunikationsnetzwerken kann unterschieden werden zwi-

Tabelle 2.2.: Performance of GAs using different types of representations for deceptive tree problems of different sizes and with different T_{opt} (arbitrary tree, MST, and star)

T_{opt}		oder 3				
		P_{succ}	fitness		t_{conv}	
			μ	σ	μ	σ
arbitrary tree	Prüfer number	0.54	0.49	(0.6)	26.0	(8.6)
	NetKey	0.78	0.23	(0.4)	23.0	(6.0)
	LB ($P_1=1$)	0.09	1.69	(0.8)	19.5	(7.6)
	LB ($P_1=20$)	0.82	0.18	(0.4)	23.3	(6.1)
	$P_1=P_2=1$	0.12	1.24	(0.6)	27.4	(8.0)
	heur. xover	0	2.63	(0.5)	8.7	(2.4)
	h. ini & xover	0	3.88	(0.1)	0.4	(0.4)

schen vermaschten und baumförmigen Netzwerken (vergleiche Abbildung 2.2). In vermaschten Netzen existieren mehrere unterschiedliche Wege für den Transport von Daten zwischen zwei Standorten. Daher muss bei vermaschten Netzwerken das Routing der Daten durch das Netzwerk gesteuert werden (wie z.B. im Internet). In der Regel existieren mehrere Wege zwischen zwei Standorten und es muss mit Hilfe von Routingmechanismen festgelegt werden, welchen Weg die Daten durch das Netz nehmen sollen (wie z.B. durch Routingtabellen für den IP-Verkehr im Internet). Diese Entscheidung wird unter anderem durch die Entfernung von Standorten als auch durch die Auslastung der unterschiedlichen Leitungen beeinflusst. Die Ermittlung von Routingtabellen ist in der Praxis recht aufwendig und es besteht die Gefahr, dass Nachrichten Umwege durch das Netz nehmen oder sogar im Kreis laufen. Vermaschte Netzwerke bieten allerdings eine erhöhte Ausfallsicherheit, da beim Vorhandensein von mehreren Wegen zwischen zwei Knoten Daten auch dann noch ausgetauscht werden können, wenn ein Teil des Netzwerkes ausfällt.

Baumförmige Netze stellen einen Sonderfall von vermaschten Kommunikationsnetzwerken dar. Hierbei existiert genau ein möglicher Weg zwischen den Knoten im Netz. Daher muss keine Entscheidung über das Routing getroffen werden und Routingmechanismen zur Steuerung des Datenflusses entfallen. Allerdings haben baumförmige Netzwerke das Problem, dass durch den Ausfall einer Leitung ein Teil der Knoten nicht mehr erreichbar sind. Baumförmige Kommunikationsnetzwerke werden oft bei einer überschaubaren Anzahl von Standorten wie

z.B. in Local Area Networks (LAN) oder Unternehmensnetzwerken („Corporate Networks“) eingesetzt. Insbesondere Unternehmen, welche eigene Netzwerke betreiben (z.B. die DATEV eG), verwenden oft baumförmige Kommunikationsnetzwerkstrukturen wegen ihrer geringeren Komplexität.

Das Problem des kostenminimalen Aufbaus von vermaschten bzw. baumförmigen Kommunikationsnetzen wurde schon sehr frühzeitig in der Literatur formalisiert und beschrieben. Beim „network design problem“ soll diejenige Netzwerkstruktur bestimmt werden, welche alle Standorte miteinander verbindet, alle Kommunikationsanforderungen zwischen den einzelnen Standorten erfüllt und gleichzeitig minimale Kosten besitzt. Aus dem allgemeinen „network design problem“ lässt sich das „optimal communication spanning tree“ (OCST)-Problem ableiten, bei welchem die Struktur des gewünschten Netzwerkes baumförmig ist. Im folgenden Abschnitt soll näher auf das OCST Problem eingegangen werden.

3. Summary, Conclusions, and Further Work

The purpose of this book is to understand the influence of representations on the performance of genetic and evolutionary algorithms. This chapter summarizes the work contained in this study and lists its major contributions.

3.1. Summary

This is the final section 3.1. We started in Chap. 1 by providing the necessary background for examining representations for GEAs. Researchers recognized early that representations have a large influence on the performance of GEAs. Consequently, after a brief introduction into representations and GEAs, we discussed how the influence of representations on problem difficulty can be measured. The chapter ended with prior guidelines for choosing high-quality representations. Most of them are mainly based on empirical observations and intuition and not on theoretical analysis.

Therefore, we presented in Chap. 2 three aspects of a theory of representations for GEAs. We investigated how the locality, scaling, and locality of an encoding influences GEA performance. The performance of GEAs is determined by the solution quality at the end of a run and the number of generations until the population is converged. Consequently, for redundant and exponentially scaled encodings, we presented population sizing models and described how the time to convergence is changed. Furthermore, we were able to demonstrate that high-locality encodings do not change the difficulty of a problem; in contrast, when using low-locality encodings, on average, the difficulty of problems changes. Therefore, easy problems become more difficult and difficult problems become easier by the use of low-locality encodings. For all three properties of encodings, the theoretical models were verified with empirical results.

3.2. Conclusions

We summarize the most important contributions of this work.

Framework for design and analysis of representations (and operators) for GEAs. The main purpose of this study was to present a framework which describes how genetic representations influence the performance of GEAs. The performance of GEAs is measured by the solution quality at the end of the run and the number of generations until the population is converged. The proposed framework allows us to analyze the influence of existing representations on GEA performance and to develop efficient new representations in a theory-guided way. Furthermore, we illustrated that the framework can also be used for the design and analysis of search operators, which are relevant for direct encodings. Based on the framework, the development of high-quality representations remains not only a matter of intuition and random search but becomes an engineering design task. Even though more work is needed, we believe that the results presented are sufficiently compelling to recommend increased use of the framework.

Redundancy, Scaling, and Locality. These are the three elements of the proposed framework of representations. We demonstrated that these three properties of representations influence GEA performance and presented theoretical models to predict how solution quality and time to convergence changes. By examining the redundancy, scaling, and locality of an encoding, we are able to predict the influence of representations on GEA performance.

The theoretical analysis shows that the redundancy of an encoding influences the supply of building blocks (BB) in the initial population. r denotes the number of genotypic BBs that represent the best phenotypic BB, and k_r denotes the order of redundancy. For synonymously redundant encodings, where all genotypes that represent the same phenotype are similar to each other, the probability of GEA failure goes either with $O(\exp(-r/2^{k_r}))$ (uniformly scaled representations) or with $O(\exp(-\sqrt{r/2^{k_r}}))$ (exponentially scaled representations). Therefore, GEA performance increases if the representation overrepresents high-quality BBs. If a representation is uniformly redundant, that means each phenotype is represented by the same number of genotypes, GEA performance remains unchanged in comparison to non-redundant encodings.

The analysis of the scaling of an encoding reveals that non-uniformly scaled

representations modify the dynamics of genetic search. If exponentially scaled representations are used, the alleles are solved serially which increases the overall time until convergence and results in problems with genetic drift but allows rough approximations of the expected optimal solution after a few generations.

We know from previous work that the high locality of an encoding is a necessary condition for efficient mutation-based search. An encoding has high locality if neighboring phenotypes correspond to neighboring genotypes. Investigating the influence of locality shows that high-locality encodings do not change the difficulty of a problem. In contrast, low-locality encodings, where phenotypic neighbors do not correspond to genotypic neighbors, change problem difficulty and make, on average, easy problems more difficult and deceptive problems easier. Therefore, to assure that an easy problem remains easy, high-locality representations are necessary.

3.3. Further Work

What are the open questions? What should be done next?

Bibliography

- McGuinness, D. L. & van Harmelen, F. (2005). OWL Web Ontology Language.
- Mustermann, M. (2009). Something that is not covered by any other annotation may be declared misc - certificates, patents and stuff like that. Currently under review.
- Padovitz, A., Loke, S. W., & Zaslavsky, A. (2008). The ecora framework: A hybrid architecture for context-oriented pervasive computing. *Pervasive Mobile Computing*, 4(2), 182–215.
- Tannenbaum, A. S. & Steen, M. V. (2007). *Distributed Systems - Principles and Paradigms*. New Jersey: Pearson Prentice Hall, 2nd edition edition.
- Tuttles, V., Schiele, G., & Becker, C. (2007). COMITY - Conflict Avoidance in Pervasive Computing Environments. In R. Meersman, Z. Tari, & P. Herrero (Eds.), *OTM Workshops (2007)*, number 4806 in Lecture Notes in Computer Science (pp. 763–772). Heidelberg: Springer.
- Welscher, C. (2001a). Indigene Völker im Gebiet Tjumen (Westsibirien): Ziele, Situation und zielorientierte Aktivitäten. Fast eine habil., Institut für Friedenssicherungsrecht und humanitäres Völkerrecht, Ruhr-Universität Bochum.
- Welscher, C. (2001b). Indigene Völker im Gebiet Tjumen (Westsibirien): Ziele, Situation und zielorientierte Aktivitäten. Master's thesis, Institut für Friedenssicherungsrecht und humanitäres Völkerrecht, Ruhr-Universität Bochum.

Appendix

A. First class of appendices

A.1. Some appendix

This is a sample appendix entry.

Eidesstattliche Erklärung

Hiermit versichere ich, dass diese Abschlussarbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurden. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Veröffentlichungen in gedruckter oder elektronischer Form sind gekennzeichnet. Sämtliche Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen.

Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen werden kann, wenn die Erklärung nicht erteilt wird.

Ort, den Datum

Martin Mustermann