

wiederverwendbaren Teillösungen und Integration zur Gesamtlösung stellt im heutigen Software Engineering den typischen Weg zur erfolgreichen Lösung komplizierter Probleme dar.

In der vorliegenden Arbeit soll die Frage betrachtet werden, ob und unter welchen Bedingungen sich bei Evolutionären Algorithmen Prozesse ausmachen lassen, die ebenfalls einer Modularisierung – also einer Entstehung modularer Strukturen – entsprechen. Dieser Sachverhalt soll hier mit dem Schlagwort *Selbstorganisierte Modularisierung* umschrieben werden.

Seit den Anfängen der Grundlagenforschung im Bereich Evolutionärer Algorithmen haben sich eine Reihe von Modellen explizit oder implizit mit dieser Problematik beschäftigt. Dabei wurden äußerst unterschiedliche Modellansätze verwendet, die von Einfachen Genetischen Algorithmen bis hin zu komplexen Klassifiziersystemen, neurogenetischen Ansätzen oder Genetischem Programmieren reichen. Wegen der Vielzahl der verschiedenen Einzelmodelle wird es zunehmend schwieriger, den Überblick über das Forschungsgebiet zu behalten.

Umgekehrt sind allgemeingültige und gleichzeitig exakte theoretische Aussagen über die Mechanismen, die *Selbstorganisierter Modularisierung* zugrundeliegen, ausgesprochen selten. Doch gerade ein solides theoretisches Fundament böte die Möglichkeit, die Vielzahl der Modelle einzuordnen, sie miteinander zu vergleichen und so eine einheitlichere Sicht auf die Mechanismen der *Selbstorganisierten Modularisierung* zu erhalten. Die Entwicklung eines solchen theoretischen Fundaments ist das Ziel der vorliegenden Arbeit.

Dieses Gesamtziel soll in drei wesentliche Teilziele untergliedert werden:

1. Es soll ein formaler Rahmen geschaffen werden, innerhalb dessen sich Modularität im Kontext Evolutionärer Algorithmen beschreiben bzw. exakt quantifizieren läßt. Um nützlich zu sein, soll der so geschaffene Modulbegriff möglichst allgemeinen Charakter haben und sich insbesondere auf bestehende Ansätze sinnvoll anwenden lassen, die die Problematik der Modularisierung thematisieren.
2. Weiterhin soll innerhalb dieses formalen Rahmens erklärt werden, wie die einfachen Mechanismen des Darwinschen Evolutionsmodells (also Selektion und Variation) die Ausprägung modularer Strukturen bewirken oder begünstigen. Zumindest für einfache Beispielfälle sollen grundlegende theoretische Betrachtungen und Voraussagen möglich sein.
3. Komplexere Fragestellungen oder Modelle sollen mittels Computersimulationen und entsprechender Visualisierung zugänglich gemacht werden. Die theoretischen Überlegungen sollen mit Hilfe von Software-Werkzeugen überprüft und die Auswirkung der Mechanismen auf die simulierten Evolutionsprozesse in komfortabler Weise beobachtet werden können.

Diese Einzel-Zielsetzungen spiegeln sich auch in der Struktur der Arbeit wider.

Zunächst soll in Kapitel 2 aufgezeigt werden, daß sich ausgehend von der biologischen Motivation zum Teil sehr unterschiedliche Typen Evolutionärer Algorithmen entwickelt haben und wie sich diese unter verschiedenen Aspekten kategorisieren lassen. Aufbauend auf diesen Betrachtungen sollen zwei verschiedene Metamodelle Evolutionärer Algorithmen betrachtet werden, die verschiedene Gesichtspunkte in unterschiedlicher Weise modellieren und die gemeinsam die formale Grundlage für die folgenden Überlegungen bilden.

---

Kapitel 3 nähert sich ausgehend von intuitiven Überlegungen dem Modularitäts-Begriff und zeigt, wie er auf den Kontext Evolutionärer Algorithmen übertragen werden kann. Durch die schrittweise Formalisierung im Rahmen der beiden Metamodelle wird dabei versucht, zunächst einen möglichst hohen Grad an Allgemeinheit zu erreichen und danach ein Maß zu entwickeln, das den späteren Betrachtungen zu Modularisierungsprozessen gerecht wird.

In Kapitel 4 wird dann demonstriert, wie sich bereits bestehende, konkrete Modellansätze zur Problematik der Modularisierung mit Hilfe der Meta-Modellierung vereinfacht darstellen lassen, und wie dann die in Kapitel 3 eingeführten Konzepte von Modularität auf diese Ansätze angewendet werden können. Dies geschieht anhand einfacher, exemplarischer Überlegungen, die gleichzeitig jedoch illustrieren, wie kompliziert allgemeine Betrachtungen zu Modularisierungs-Prozessen in den bestehenden Modellen werden können.

Um trotzdem eine abstrakte Sicht auf solche Prozesse gewinnen zu können, wird im Anschluß ein stark vereinfachtes informatisches Metamodell entwickelt, das eine Reihe von wichtigen gemeinsamen Eigenschaften der bestehenden Modelle in sich trägt.

Die Frage nach den Möglichkeiten, wie die einfachen Mechanismen der Evolution eine Ausprägung von Modulen bewirken können, wird dann in Kapitel 5 thematisiert. Mit Hilfe der in Kapitel 3 eingeführten Formalismen lassen sich für einfache Spezialfälle Zusammenhänge finden, die eine theoretische Begründung für Prozesse *Selbstorganisierter Modularisierung* liefern. Simulationsexperimente erlauben zum einen die Überprüfung der Voraussagen und zum anderen einen interessanten Einblick in das theoretisch nicht oder nur schwer zugängliche Gebiet der längerfristigen Dynamik.

Kapitel 6 widmet sich der software-technischen Umsetzung der Konzepte im allgemeinen und des entwickelten Metamodells im besonderen. Dabei werden verschiedene, weitgehend plattformunabhängige Komponenten beschrieben, ebenso die im Rahmen der Arbeit erstellte Simulations- und Visualisierungsumgebung, die sowohl interaktive Versuche als auch systematische Simulationsexperimente erlaubt.

Eine Zusammenfassung der Ergebnisse und ein Ausblick auf mögliche Weiterentwicklungen in Kapitel 7 runden die Arbeit ab.



---

## Kapitel 2

# Evolutionäre Algorithmen

*In diesem Kapitel soll ein kurzer Einblick in den Forschungsbereich Evolutionärer Algorithmen gegeben werden, der den wissenschaftlichen Kontext der vorliegenden Arbeit darstellt. Zunächst werden in Abschnitt 2.1 einige wichtige Aspekte der biologischen Grundlagen dargestellt, die als Motivation und Ausgangspunkt für diese Algorithmen dienen. Danach wird in Abschnitt 2.2 auf die verschiedenen Möglichkeiten der Modellierung von Evolution im Rahmen Evolutionärer Algorithmen eingegangen. Neben der Schaffung einer Übersicht über das Themengebiet soll diese Betrachtung die Schwierigkeit verdeutlichen, für die Vielfalt der Modellierungs-Ansätze ein übergreifendes Konzept für Selbstorganisierte Modularisierung zu finden. Den Ausgangspunkt für das formale Konzept, innerhalb dessen diese Schwierigkeit im Rahmen dieser Arbeit überwunden wird, bilden zwei aus den 1990er Jahren stammende Metamodelle, die abstrakte Modelle Evolutionärer Algorithmen darstellen. Diese werden in Abschnitt 2.3 vorgestellt und charakterisiert. In Abschnitt 2.4 werden die Ergebnisse des Kapitels kurz zusammengefaßt.*

### 2.1 Biologische Motivation

Motivation und Ausgangspunkt Evolutionärer Algorithmen ist die biologische Evolutionstheorie, die heute im Rahmen der Biologie in ihren Grundzügen weitgehend anerkannt ist. Der Beitrag der biologischen Evolutionstheorie beschränkt sich dabei durchaus nicht auf eine vertraut erscheinende Nomenklatur oder auf inspirierende Metaphern. Vielmehr liefert diese Theorie tatsächlich die grundlegenden Kern-Konzepte, die die Dynamik und Funktion der darauf basierenden abstrakten Modelle bestimmen. Aus diesem Grund erscheint es lohnenswert, bestimmte Aspekte der biologischen Evolutionstheorie kurz zu umreißen.

### 2.1.1 Darwins Theorie zur Evolution und ihre Weiterentwicklung

Überlegungen zur Entstehung und Entwicklung des Lebens sind wahrscheinlich annähernd so alt wie die Menschheit. Manche davon haben in vielfältigen Schöpfungsmythen die Zeiten überdauert. Auch in der antiken Philosophie sind bereits Bestrebungen zu finden, Aussagen über Entwicklungsprozesse zu machen. Erste naturwissenschaftliche Betrachtungen im heutigen Sinne, die für die Entwicklung der Evolutionstheorie von Bedeutung waren, wurden seit dem 18. Jahrhundert angestellt.

Die wohl bekannteste unter den Evolutionstheorien vor Darwin ist die Theorie von Jean Baptiste de Lamarck (1744-1829) [de Lamarck, 1809]. Als Mechanismen für die Evolution sah Lamarck den Gebrauch bzw. Nichtgebrauch von Organen durch die Individuen, der – so Lamarck – nicht nur Auswirkungen auf die Organe des jeweiligen Individuums haben sollte, sondern sich auch – durch Vererbung – auf die entsprechenden Organe der Nachkommen auswirken sollte. Eine Weitergabe von erworbenen Eigenschaften in dieser naiven Form konnte durch die moderne Biologie nicht bestätigt werden, so daß die Lamarcksche Theorie der Evolution heute zumindest in einem Teil ihrer Aussagen als obsolet angesehen wird.

#### Darwins Evolutionstheorie

Die Grundlage der allermeisten der heute üblichen Evolutionstheorien ist die Evolutionstheorie des britischen Naturforschers Charles Darwin (1809-1882), die zumindest in Teilen auch unabhängig von Alfred R. Wallace (1823-1913) aufgestellt wurde. Die Theorie beruht zum einen auf Beobachtungen bei wissenschaftlichen Expeditionen, zum anderen aber auch auf der Analogie zwischen natürlichen Lebensbedingungen von Lebewesen und den von Tierzüchtern praktizierten Ausleseverfahren.

Die Theorie, die Darwin in seinem berühmt gewordenen Werk „The Origin of Species“ [Darwin, 1859] darlegte, läßt sich durch die Betrachtung einiger zentraler Beobachtungen und einfachen Schlußfolgerungen skizzieren (hier zitiert nach [Campbell, 1997]):

**Beobachtung 1** Alle Arten weisen ein derart hohes Fortpflanzungspotential auf, daß ihre Populationsgröße exponentiell zunehmen würde, wenn alle Individuen, die geboren werden, sich erfolgreich fortpflanzen würden.

**Beobachtung 2** Die meisten Populationen sind normalerweise mit Ausnahme saisonaler Schwankungen in ihrer Größe stabil.

**Beobachtung 3** Die natürlichen Ressourcen sind beschränkt.

**Folgerung 1** Die Produktion von mehr Nachkommen, als die Umwelt tragen kann, führt unter den Individuen einer Population zum Kampf ums Überleben, wobei in jeder Generation nur ein Bruchteil des Nachwuchses überlebt.

**Beobachtung 4** Die Individuen einer Population variieren enorm in ihren Merkmalen; keine zwei Individuen sind exakt gleich.

**Beobachtung 5** Ein Großteil dieser Variabilität [sic!] ist erblich.

**Folgerung 2** Das Überleben im Existenzkampf beruht nicht auf Zufall, sondern hängt unter anderem von den Erbanlagen der überlebenden Individuen ab. Die durch ihre ererbten Merkmale am besten an die Umwelt angepaßten Individuen hinterlassen wahrscheinlich mehr Nachkommen als weniger gut angepaßte.

**Folgerung 3** Die ungleichen Überlebens- und Fortpflanzungsfähigkeiten von Individuen führen zu einem graduellen Wandel in einer Population, wobei sich vorteilhafte Merkmale im Laufe der Generationen anhäufen.

Eine Vielzahl von biologischen Beobachtungen und paläontologischen Funden scheinen die Darwinsche Theorie der Evolution zu stützen, die den Kern der heutigen Standardtheorie der biologischen Evolution bildet. Dem Wissensstand der damaligen Zeit entsprechend betrafen Darwins Aussagen jedoch nur die Ebene des sichtbaren Erscheinungsbildes von Individuen; über die genetischen Grundlagen der Vererbung war zu seiner Zeit kaum etwas bekannt.

### Die Weiterentwicklung von Darwins Theorie

Einen ersten Zugang zu den Mechanismen der Vererbung schuf ein Zeitgenosse Darwins, der Augustinermönch Johann Gregor Mendel (1822-1884). Dieser entwickelte weitgehend unbeachtet von der Öffentlichkeit aufgrund von Experimenten und theoretischen Überlegungen die „Mendelschen Gesetze“ der Genetik, die ausgehend von Beobachtungen im Erscheinungsbild von Individuen wichtige Rückschlüsse auf die Mechanismen der Vererbung zulassen. Nach Mendels Tod gerieten seine Arbeiten für einige Zeit in Vergessenheit und wurden erst zu Beginn des 20. Jahrhunderts wiederentdeckt.

Zu diesem Zeitpunkt schien Mendels Vorstellung von diskreten Erbanlagen zunächst im Widerspruch mit der Darwinschen Annahme zu stehen, daß Merkmale sich allmählich verändern.

Erst mit der Weiterentwicklung des Forschungsgebietes der Genetik konnte dieser scheinbare Widerspruch aufgelöst werden. In den 1930er und 1940er Jahren des 20. Jahrhunderts bildete sich der Forschungszweig der Populationsgenetik heraus. Verschiedene Arbeiten, etwa von Ronald A. Fisher (1890-1962), John B.S. Haldane (1892-1964), Theodosius Dobzhansky (1900-1975), Julian Huxley (1887-1975) und Ernst Mayr (geb. 1904), führten zur sogenannten *Synthetischen Theorie*, die populationsgenetische und evolutionsbiologische Erkenntnisse vereinigte [Futuyma, 1990].

#### 2.1.2 Eine moderne Sicht auf die genetischen Grundlagen der Variationsmechanismen

Durch die stark erweiterten Erkenntnisse im Bereich der Molekularbiologie bzw. der Molekulargenetik ist – ungeachtet von Differenzen in bestimmten Details und Sichtweisen – ein in vielen Punkten einheitliches Bild der Evolutionsbiologie entstanden. Die von Darwin postulierten bzw. phänomenologisch beobachteten Vorgänge der Vererbung und der dabei auftretenden Variationen konnten dabei auf eine molekulargenetische Grundlage gestellt werden. Dieses konkrete Bild hat die Entwicklungen in weiten Bereichen der evolutionären Algorithmen entscheidend geprägt.

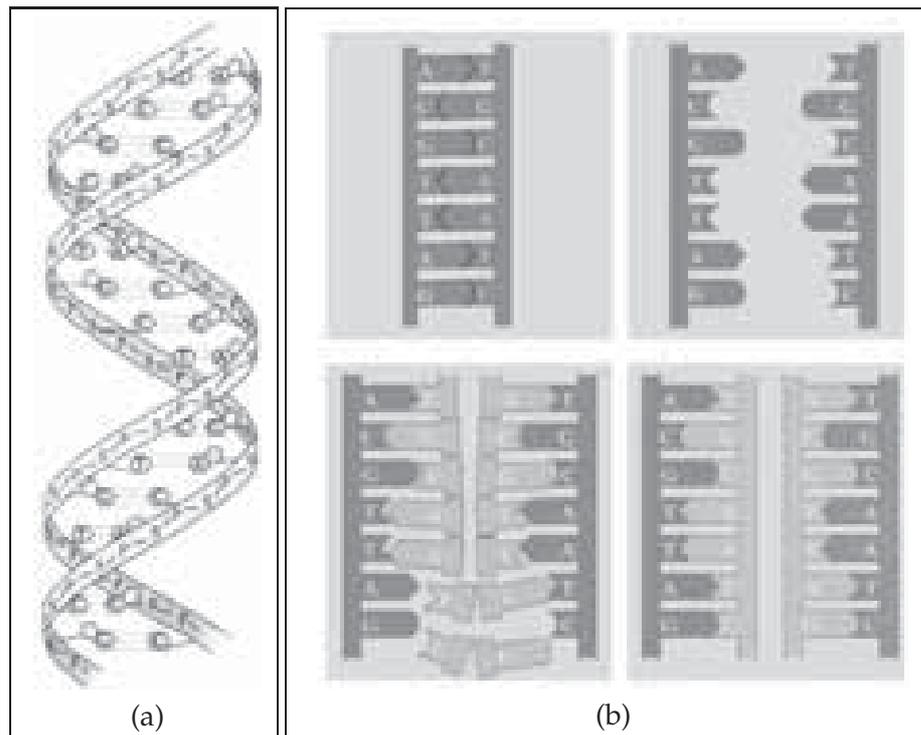
#### Die DNA als Träger der Erbinformation

Als Träger der Erbinformation wurde seit Mitte der 1940er Jahre die DNA durch verschiedene Experimente identifiziert (entgegen der vorher vorherrschenden Meinung, Träger der Erbinformation seien Proteine). Ausnahmen von der DNA-Codierung sind bestimmte Sorten von Viren, bei denen die Erbinformation in Form von RNA-Molekülen gespeichert ist, wobei die Codierung hierbei auf einer abstrakteren Ebene sehr ähnlich ist. Die Aufklärung der Molekülstruktur von DNA gelang 1953 durch den Biologen James Watson und den Biophysiker Francis Crick und gilt als einer der wichtigsten Meilensteine der modernen Biologie.

DNA-Moleküle bestehen aus einer Doppelhelix mit zwei antiparallelen Molekülsträngen, die über Brücken miteinander verbunden sind. Jede dieser Brücken setzt sich

aus jeweils zwei verschiedenen Kernbasen (Nukleotiden) zusammen, so daß man dabei von Basenpaaren spricht (s. Abb. 2.1a). Insgesamt können vier unterschiedliche Kernbasen unterschieden werden: Adenin, Guanin, Cytosin und Thymin (abgekürzt: A,G,C,T). Die Basenpaare bestehen immer entweder aus einer Adenin- und einer Thymin-Base oder aber aus einer Cytosin- und einer Guanin-Base. Diese Komplementarität ermöglicht es, daß DNA-Moleküle in vergleichsweise einfacher Art und Weise kopiert werden können, weil bei einer Spaltung des Doppelstrangs in zwei Teilstränge der jeweils fehlende immer in eindeutiger Weise ergänzt werden kann. (s. Abb. 2.1b).

In einer abstrakten Betrachtungsweise ist also die „Sprache“, in der die Erbinformation codiert ist, aus einem Alphabet von vier Zeichen aufgebaut, nämlich genau A,G,C und T.



**Abb. 2.1:** **a)** Darstellung eines Ausschnitts eines als Doppelhelix gewundenen DNA-Moleküls. Die chemischen Strukturen sind dabei deutlich vereinfacht dargestellt. **b)** Schematische Darstellung des Kopiermechanismus. Ein DNA-Doppelstrang kann sich aufspalten und freie Nukleotide können sich in komplementärer Weise anlagern. Auf diese Weise entstehen zwei identische Stränge. (Beide Abbildungen nach [Campbell, 1997], verändert.)

Bestimmte Abschnitte auf der DNA, die z.B. die Struktur eines bestimmten Eiweißes codieren, werden als „Gene“ bezeichnet. Diese Abschnitte können mehrere hundert Basenpaare beinhalten. Verschiedene Konfigurationen ein und desselben Gens heißen „Allele“. Die Gesamtheit der Erbinformation wird als „Genom“ bezeichnet. Bei den meisten höheren Lebewesen besteht die Erbinformation nicht aus einem einzigen DNA-Doppelstrang, sondern die Erbinformation ist in mehreren Chromosomen untergebracht.

Bei vielen dieser Lebewesen liegen in jedem Individuum jeweils mehrere Chromosomen vor, die in redundanter Weise für die Codierung der gleichen Information zuständig sind. Bei allen Wirbeltieren, aber zum Beispiel auch bei den von Mendel beobachteten Erbsenpflanzen sind dies jeweils zwei Chromosomen. Man spricht hier vom Vorhandensein eines „doppelten Chromosomensatzes“ und von „diploiden“ Individuen; Individuen mit nur einem Chromosomensatz heißen „haploid“. Man findet jedoch auch Lebewesen mit

noch höherer Redundanz im Chromosomensatz: Dazu gehören beispielsweise Pflanzen, die in Gebirgsregionen mit hoher UV-Bestrahlung vorkommen. Teilweise kommt es sogar vor, daß innerhalb ein und derselben Art Unterschiede in der Redundanz des Chromosomensatzes vorkommen: so sind die Königinnen und Arbeiterinnen in Bienenvölkern diploid, während die Drohnen nur haploid sind.

Die genetischen Strukturen der Individuen sind die Basis, auf der sich die genetische Vererbung vollzieht und auf der auch genetische Varianz entstehen kann.

### Variation durch Mutation und Vermischung von genetischem Material

Zwei vom Ansatz her verschiedene Mechanismen spielen bei der Entstehung von genetischer Varianz eine entscheidende Bedeutung: Das eine sind zufällige Veränderungen von genetischem Material, sogenannte *Mutationen*. Das andere ist die Vermischung von genetischem Material, die vor allem durch die Mechanismen *Rekombination* und *Crossing over* realisiert wird.

**Mutationen** sind Veränderungen an der DNA, die durch verschiedene Einflüsse wie etwa Radioaktivität, UV-Strahlung oder chemische Einwirkungen zustandekommen können. Im einfachsten Fall wird durch Kopierfehler eine Kernbase durch eine andere ausgetauscht (Punktmutationen), in anderen Fällen werden Kernbasen eingefügt (Insertionen) oder weggelassen (Deletionen).

**Vermischung von genetischem Material** kommt durch zwei Mechanismen zustande, die sich auf unterschiedlichen Ebenen abspielen. Unter *Rekombination* wird dabei meist eine Vermischung von Erbgut auf der Ebene von Chromosomensätzen verstanden: bei der Bildung von Keimzellen (also Samenzellen und Eizellen) wird jeweils nur ein einfacher Chromosomensatz übernommen, (d.h. die Keimzellen sind haploid). Erst bei der Verschmelzung von Samen- und Eizelle bildet sich wieder ein Individuum mit komplettem, d.h. doppeltem Chromosomensatz).

Dagegen bezeichnet der Begriff *Crossing over* die Vermischung von Erbgut innerhalb eines einzelnen Chromosoms. Eine detaillierte Darstellung findet sich in vielen Standard-Lehrbüchern der Biologie wie z.B. [Campbell, 1997].

### Genotyp und Phänotyp

Die in der DNA eines Lebewesens codierte Information (sein *Genotyp*) kann als Bauplan für das Lebewesen aufgefaßt werden. Die „Übersetzung“ dieses Bauplans in Strukturen und Funktionen innerhalb des entsprechenden Lebewesens, den sogenannten *Phänotyp*, ist ausgesprochen kompliziert und bis heute zu großen Teilen ein offenes Forschungsfeld. Ein zentraler Mechanismus bei der Umsetzung des Genotyps in den Phänotyp ist die Übersetzung von DNA-Information in Abfolgen von Aminosäuren in Eiweißketten (s. Abb. 2.2). Die Reihenfolge der Aminosäuren wird durch die Abfolge der Nukleotide in Teilen der DNA bestimmt. Jeweils drei Nukleotide codieren dabei (zum Teil redundant) eine von 20 möglichen Aminosäuren. Die Eiweiße selbst können durch Faltung unterschiedlichste Formen annehmen und vielfältige Aufgaben im biochemischen Gefüge von Organismen erfüllen. Insbesondere können sie als Enzyme wirken, die als „Biokatalysatoren“ die für die Lebensvorgänge notwendigen chemischen Prozesse in Gang setzen. Die Vorstellung der DNA als ein „Bauplan“ ist insofern etwas irreführend, als technische Baupläne typischerweise nur bei der Konstruktion von Maschinen verwendet werden