

# Algorithmen auf Sequenzen (IN5022)

Title	Algorithms on Sequences	
Typ	Vorlesung mit Übungen	
Credits	9	
Lehrform/SWS	4V + 2Ü	
Sprache	Deutsch	
Modulniveau	Master	
Arbeitsaufwand	Präsenzstunden	90 Stunden
	Eigenstudium	180 Stunden
	Gesamtaufwand	270 Stunden
Angestrebte Lernergebnisse	<p>Die Teilnehmer sind in der Lage, Problemstellungen auf Sequenzen für einen algorithmischen Zugang zu modellieren, damit sie einem automatisierten Lösungsverfahren zugänglich sind. Beispiele sind die Analyse bzgl. der Reversionsdistanz u.Ä., Mustererkennung in Sequenzdaten, Textindexierung, speicherplatzsparende Darstellung, etc. für biologische Sequenzen, wie z.B. DNA- und Aminosäuresequenzen. Studierende können Algorithmen für deren Lösung entwerfen und diese im Hinblick auf deren Effektivität (Korrektheit) und Effizienz (bzgl. Laufzeit und insbesondere auch Speicherplatzverbrauch) analysieren und die algorithmische Komplexität (bzgl. der Komplexitätsklassen P, NP, NPC, etc.) des zugehörigen Problems bestimmen und einordnen.</p>	
Intended Learning Outcomes	<p>For problems based on sequences participants are able to construct appropriate models to make the problems accessible for algorithmic solutions. Examples are the analysis of reversal and similar distances, pattern recognition in sequences, text indexing, succinct representations, etc. for biological DNA and amino acid sequences. Students are able to design and to analyze (with respect to time and particularly space complexity as well as correctness) algorithms for the proposed models and to evaluate their algorithmic complexity (with respect to P, NP, NPC, etc.).</p>	
Inhalt	<p>Dieses Modul umfasst die grundlegenden Problemstellungen der Bioinformatik, die auf Sequenzen beruhen, das Erlernen und Anwenden allgemeiner und spezifischer Lösungsmethoden hierfür, das Analysieren der algorithmischen Komplexität und das Entwerfen von Algorithmen für</p>	

	<p>solche Problemstellungen. Themen sind die Ermittlung optimal bewerteter Teilfolgen, Konstruktion und Einsatz von Suffix-Bäumen und Suffix Arrays und u.a. Verfahren zur Textindexierung, spezielle Methoden zur Mustererkennung in Sequenzen, wie Repeats u.Ä., besonders speicherplatzeffiziente Methoden (wie Suffix-Arrays, Rank-Select, Burrows-Wheeler-Transformation, etc.), mathematische Modellierung und Methoden zur Lösung der Bestimmung der Reversionsdistanz und ähnlicher Maße.</p>
<p>Contents</p>	<p>The module treats bioinformatics problems based on sequences, the design of general and specific techniques for these problems, techniques to analyze the algorithmic complexity, and the design of algorithms for such problems. Topics are computing optimal scoring subsequences, construction and applications of suffix trees and suffix arrays and other text indexing methods, specific algorithms for pattern recognition in sequences such as repeats, succinct data structures (such as suffix arrays, rank-select, Burrows-Wheeler transformation), mathematical modeling and algorithms for the computation of the reversal and similar distance.</p>
<p>Prüfung</p>	<p>Prüfungsleistung (benotet): -Klausur: 120 min</p> <p>Wiederholungsklausur zu Ende des Semesters. Details werden zu Beginn des Moduls bekannt gegeben.</p> <p>In der Klausur weisen die Studierenden nach, dass sie Problemstellungen auf Sequenzen (insbesondere biologische, wie z.B. DNA- und Aminosäuresequenzen) für einen algorithmischen Zugang modellieren können, indem sie Modelle und Algorithmen auf verwandte Probleme erweitern bzw. übertragen sowie bzgl. ihrer Komplexität analysieren.</p> <p>Konkret werden in der Klausur 4-7 Aufgaben bearbeitet, die eine eigenständige Anwendung der algorithmischen Konzepte und Modelle aus der Vorlesung zur Lösung einer anspruchsvollen Problemstellung erfordern (Anwendung oder Entwurf eines Algorithmus beispielsweise zur Sequenzanalyse, Mustererkennung in Sequenzen,</p>

	<p>Textindexierung mit Suffix-Bäumen, -Arrays o.Ä., Bestimmung genomischer Distanzen wie Reversionsdistanz), die Skizzierung einer Beweisidee (z.B. polynomielle Reduktion auf ein NPC-Problem, Nachweise polynomieller Laufzeit oder Korrektheit) oder die Analyse der Zeit- und insbesondere Platzkomplexität eines Algorithmus (z.B. im Bereich Sequenzanalyse, Textindexierung, Mustererkennung, besonders platzeffizienter Methoden, genomischer Distanzen).</p>
<p>Examination</p>	<p>Examination requirements (graded): -written exam : 120 min</p> <p>A makeup exam will be offered at the end of the semester, details will be announced at the beginning of the course.</p> <p>Within the written exam, students demonstrate that they are able to model given problems on sequences (particularly biological sequences such as DNA and amino acid sequences) in such a way that algorithmic approaches are applicable. For this purpose, they transfer and extend models and algorithms to similar problems and to analyze their complexity. The written exam contains 4 to 7 assignments, which require independent application of algorithmic concepts and models presented in the lecture to solve demanding problems (application or design of algorithms ,e.g. for sequence analysis, pattern recognition in sequences, text indexing using suffix trees and arrays as well as similar concepts, computation of genomic distances such as reversal distance, etc.), the sketch of a proof (e.g., polynomial reduction to an NP-complete problem, proof of a polynomial runtime, proof of correctness of an algorithm) or the analysis of the time and in particular space complexity of an algorithm (e.g., in the field of sequence analysis, text indexing, pattern recognition, succinct data structures, genomic distances, etc.).</p>
<p>Literatur</p>	<p>S. Aluru (Ed.): Handbook of Computational Molecular Biology, Chapman and Hall/CRC, 2006. D. Gusfield: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology, Cambridge University Press, 1997. G. Fertin, A. Labarre, I. Rusu, E. Tannier, S. Vialette: Combinatorics of Genome Rearrangements, MIT</p>

	<p>Press, 2009.</p> <p>V. Heun: Algorithmen auf Sequenzen, Skriptum Institut für Informatik der LMU</p>
Literature	<p>S. Aluru (Ed.): Handbook of Computational Molecular Biology, Chapman and Hall/CRC, 2006.</p> <p>D. Gusfield: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology, Cambridge University Press, 1997.</p> <p>G. Fertin, A. Labarre, I. Rusu, E. Tannier, S. Vialette: Combinatorics of Genome Rearrangements, MIT Press, 2009.</p> <p>V. Heun: Algorithmen auf Sequenzen, Class Notes, Institute for Informatics, LMU</p>
Medienformen	Tafelanschrieb, Folienpräsentation
Media	Blackboard, slide show
Lehr- und Lernmethode	<p>Vorlesung, Tutorübung, Aufgaben zum Selbststudium.</p> <p>Das Modul besteht aus einer Vorlesung sowie Übungen in kleinen Gruppen. In den Hausaufgaben, die freiwillig abzugeben sind, wird das Verständnis der Konzepte und Algorithmen, die in der Vorlesung vorgestellt werden, anhand konkreter Daten und Beispiele vertieft. Die Studierenden entwerfen mithilfe der in der Vorlesung vorgestellten Konzepte und Modelle neue Algorithmen für neue Probleme und analysieren diese. In den Hausaufgaben werden selbständig anspruchsvolle Übungsaufgaben bearbeitet, die ähnlich wie die Klausuraufgaben sind (siehe oben) und daher zur Vorbereitung darauf dienen.</p> <p>In den Übungen werden mögliche Lösungsansätze der Aufgaben zum Selbststudium diskutiert.</p>
Teaching and Learning Methods	<p>Lecture, tutorial, assignments for individual study. The module consists of a lecture and in addition exercises in small groups. Within the assignments (submission is optional), concepts and algorithms (presented in the lecture) will be applied to real data and examples to deepen the understanding. The students develop new algorithms based on the presented concepts and models to related problems and analyze these algorithms. The assignments consist of demanding problems similar to the assignments in the written exam (for details see above) and serve as a preparation for the written exam. Within the tutorials possible approaches for solutions of the assignments will be discussed.</p>

Turnus	unregelmäßig
Modulverantwortlicher	Prof. Dr. Volker Heun
Dozenten	Prof. Dr. Volker Heun