Das Thema Elena's Präsentation war "Multiple Stringsuche: Verfahren von Aho-Corasick". Zuerst hat sie eine Übersicht dazu gegeben, das sind die Motivation, Pattern Matching, Knurt-Morris-Pratt-Algorithmus, Aho-Corasick-Algorithmus mit der Übergangsfunktion, Fehlerfunktion und Ausgabefunktion, ihr Laufzeit und Verwendung.
Bei dem Pattern Matching versteht man eine Mustererkennung beziehungsweise eine Suche nach Schlüsselwörtern. Heutzutage wird es in vielen Bereiche der Informatik immer wichtiger. Der naivste Ansatz des Pattern Matching ist der Suche nach Muster x der Länge m im Text t der Länge n. Die Laufzeit beträgt O(n*m).
Im Gegensatz zum naiven Algorithmus speichert der Knurt-Morris-Pratt-Algorithmus Informationen über bereits gewonnene Erkenntnisse bei der Zeichensuche ab. Hier werden die Informationen in einer Sprungtabelle gespeichert. Auch es ist nicht notwendig die Suche nach einer fehlenden Übereinstimmung von vorne zu beginnen und die Laufzeit ist deutlich geringer O(n+m).
Das Ziel wird durch Beantwortung auf solcher Frage definiert, nämlich was ist aber wenn nach mehreren Schlüsselwörter gesucht werden soll. Wenn O(m+n) bei einem Schlüsselwort, dann wenn O(m+n) bei einem Schlüsselwort. Also der Text wird für jedes Schlüsselwort erneut durchgelaufen. Deswegen das Ziel ist den Faktor k zu eliminieren und das Suchen mehrerer Schlüsselwörter gleichzeitig bei einem Suchdurchgang zu ermöglichen.
Die Idee, die dahinter steckt ist die Mischung aus Knuth-Morris-Pratt mit endlichen Automaten aber dazu wird nötig: Keyword-Baum (engl. Trie) für Patternmenge P.
Nächster Algorithmus ist Aho-Corasick, der wurde von zwei kanadischen Informatiker Alfred V. Aho und Margeret J. Corasick entwickelt. Er ermöglicht gleichzeitige Suche nach mehreren Schlüsselwörtern und konstruiert einen deterministischen endlichen Automat und wird durch 6-Tupel der Form (Q, Σ, g, f, out, q0) beschrieben. Die Übergangsfunktion g stellt eine Repräsentation der Schlüsselwörter in einer Baumstruktur dar.
Dann hat Elena ein Beispiel zu einem Trie gegeben - P = {he, she, his, hers}, nämlich die Konstruktion und wie wird eine Suche da funktioniert.
Als Nächstes kam die Fehlerfunktion, welche definiert die sogenannten Fehler-Links. Der Link verweist von Zustandsknoten v auf einen Knoten w im Baum. Die Fehler-Links der Knoten der Tiefe 1 führen grundsätzlich zur Wurzel zurück und die Berechnung der Links der anderen Knoten wird unter der Beachtung der Reihenfolge der Tiefen durchgeführt.
Ausgabefunktion gibt zu jedem Zustand eine Menge von Schlüsselwörtern an, die in diesem Zustand gefunden wurden und die Laufzeit wird nach folgende Weise definiert - O(m+n+k), wo m - Gesamtzahl der Zeichen alles Schlüsselwörter, n - die Länge des Textes, k - die Anzahl der vorkommenden Schlüsselwörter im Text.
Generell die Verwendung des Aho-Corasick Algorithmus kommt meistens in der Bildverarbeitung, besonders bei Bildvergleich, sowie in der Erkennung von Virenmuster, die sich in unseren Netzwerken befinden vor. Auch oft in der Bioinformatik, bei der Durchsuchung der menschlichen DNA und in vielen anderen Suchen in der Medizin, da dort vor allem mir sehr großen Datenmengen gearbeitet wird.