

Ausarbeitung im Fach Data Mining  
Hochschule Fulda  
Angewandte Informatik, Semester 5

## Data Mining am Beispiel des Spionageprogramms Echelon



Julius Helm

eMail: [julius.helm@informatik.hs-fulda.de](mailto:julius.helm@informatik.hs-fulda.de)

Hiermit bestätige ich, dass diese Ausarbeitung selbständig und ohne fremde Hilfsmittel von mir angefertigt wurde.

---

Julius Helm

## Data Mining am Beispiel des Spionageprogramms Echelon

1. Einleitung
2. Was ist Echelon
  1. Historisches
  2. Betreiber
  3. Umfang
  4. Ziele
  5. Grenzen
  6. Interessantes
3. Datensammlung
  1. Welche Daten werden gesammelt?
  2. Woher kommen die Daten
  3. Art der Daten
4. Datenaufbereitung
  1. Filterung
  2. Verteilung
  3. OCR, Spracherkennung, Übersetzung
  4. Entschlüsselung
5. Software
  1. Memex
  2. Intelink
6. Data Mining Techniken
  1. Indexierung
  2. Metadaten
  3. Text Mining
  4. Natural Language Processing
  5. Pattern Matching
  6. Mustererkennung
  7. Verborgene Markow Modelle
  8. Künstliche Intelligenz
  9. Künstliche Neuronale Netze
7. Schlussfolgerung
8. Quellenangaben

## 1. Einleitung

Dass die US-Regierung mit Echelon ein umfassendes, weltweites Spionageprogramm betreibt ist lange kein Geheimnis mehr, das Programm wird sogar offiziell bestätigt<sup>1</sup>. Verborgen bleiben der Öffentlichkeit allerdings Informationen darüber, in welchem Umfang und wie die Spionage erfolgt. Auch genaue Informationen wozu das System eingesetzt wird sind nicht bekannt. Informationen hierzu lassen sich lediglich erahnen. Zum Beispiel aus Patentanträgen<sup>2 3</sup>, die von den Betreibern des Systems, kooperierenden Firmen, oder Hochschulen gestellt wurden, sowie auf Grund von Auskünften ehemaliger Mitarbeiter<sup>4</sup>, und aus aktuellen Auftragsvergabeverfahren. Besonders im Internet finden sich viele solcher Informationen, allerdings sind diese weit verstreut und selten vollständig. Diese Ausarbeitung ist ein Versuch diese Informationen zusammenzutragen und zu strukturieren, um schließlich eine Vermutung zu ermöglichen, wie leistungsfähig Echelon wirklich ist, und welche Methoden des Data Mining<sup>5</sup> angewendet werden. Weiter soll ein Ausblick auf künftige Möglichkeiten des Systems gegeben werden.

Der Autor möchte nochmals vorweg betonen, dass es sich bei den dargestellten Sachverhalten größtenteils um ungesicherte Informationen handelt, da das Spionageprogramm der Geheimhaltung unterliegt und Informationen nur selten offiziell bestätigt werden.

## 2. Was ist Echelon<sup>6</sup>

### Historisches<sup>7</sup>

Echelon ist ein Produkt des Kalten Krieges. Sowohl die Staaten der westlichen, als auch die der östlichen Hemisphäre unterhielten leistungsfähige Abhörmaschinerien um taktische Informationen des jeweiligen Gegenparts in Erfahrung zu bringen. Aber nicht nur militärische Informationen<sup>8</sup> waren von Interesse, auch Nachrichten über Politik, die wirtschaftliche Lage, sowie das Gewinnen propagandatauglicher Negativschlagzeilen waren gerne gesehene Nebenprodukte.

---

1 Vgl. 3) Kai Billen (): „ECHELON – Das globale Abhörnetzwerk“

2 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

3 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“ (Kapitel Software)

4 Vgl. 8) Nicky Hager (2000): „Wie ich Echelon erforscht habe“

5 Vgl. 1) Wikipedia (2007): „Data Mining“

6 Vgl. 2) Wikipedia (2007): „ECHELON“

7 Vgl. 5) Duncan Campbell (2000): „Inside Echelon“ (Kap. Die Anfänge)

8 Vgl. 2) Wikipedia (2007): „ECHELON“

Bis vor wenigen Jahrzehnten beschränkte sich die Informationsbeschaffung auf das Abhören, Dechiffrieren und Erfassen von Telefonaten, Radiomeldungen und der Satelitenkommunikation. Seit dem Aufkommen digitaler Kommunikationsmittel haben sich die Anforderungen an die Spionagesysteme jedoch stark geändert. Die Analyse des digitalen oder digitalisierten Datensammelsuriums stellt die hohe Kunst des Data Mining dar. Aus gesammelten Informationen lassen sich über Querverweise zu ähnlichen oder gleichen Informationen neue Informationen generieren, die den beiden ursprünglich vorliegenden Nachrichten noch gar nicht zu entnehmen sind. Auch die Speicherung und Filterung digitaler Kommunikation ist um ein Vielfaches einfacher und effektiver. Dass die Weiterentwicklung der Computertechnologie ungeahnte Möglichkeiten in der Dechiffrierung eröffnet braucht wohl kaum erwähnt zu werden.

Eine weitere große Wende für das Echelonprogramm stellte der Fall des eisernen Vorhangs dar. War das Spionageprogramm offiziell doch primär zur Gewinnung von Informationen über den politischen Gegner eingerichtet worden, müsste nun, da dieser per se nicht mehr vorhanden war, das ganze Programm eingestampft werden. Allerdings diene gerüchteweise das System ja nicht nur hierfür, sondern auch – was offiziell nie bestätigt wird – zur Wirtschaftsspionage, der Auffindung von Kriminellen und der Überwachung offiziell befreundeter Staaten. So haben sich die Ziele von Echelon in den letzten knapp 20 Jahren deutlich geändert.

Betreiber<sup>9 10 11 12 13 14 15 16 17</sup>

Betrieben wird das Echelon-Netzwerk vom 1948 gegründeten UKUSA Zusammenschluss und befreundeten Geheimdiensten. Wie der Name vermuten lässt stecken hinter dem Zusammenschluss zuerst einmal die Geheimdienste des UK (United Kindom = Großbritannien) – der SIS (Secret Intelligence Service), sowie der Geheimdienst der USA – die NSA (National Security Agency). Letzter ist der größte in diesem Zusammenschluss, und vermutlich auch der größte der Welt. Genaue Zahlen unterliegen leider der Geheimhaltung und waren daher nicht in Erfahrung zu bringen. Zum UKUSA

---

9 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

10 Vgl. 2) Wikipedia (2007): „ECHELON“

11 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

12 Vgl. 5) Duncan Campbell (2000): „Inside Echelon“

13 Vgl. 9) Ingo Ruhmann, Christiane Schulzki-Haddouti (1998): „Lauschangriff / Abhör-Dschungel“

14 Vgl. 16) Wikipedia (2007): „UKUSA“

15 Vgl. 10) unbekannt (): „NSA der große Bruder hört mit“

16 Vgl. 11) unbekannt (): „Echelon“

17 Vgl. 40) Institut für Geschichte, Salzburg (1999): „Big Brother is watching you“

Zusammenschluss kommen noch hinzu die Geheimdienste der Staaten Kanada, Australien und Neuseeland. Befreundete angeschlossene Nationen sind z. B. Norwegen, Dänemark, Deutschland und die Türkei. Alle beteiligten Staaten haben Zugriff auf das System nur innerhalb ihres selbst definierten Interessenprofils; Vollzugriff auf das gesamte System hat lediglich die NSA.

### Umfang <sup>18 19</sup>

Zum Umfang des Systems können nur Vermutungen angestellt werden. Offizielle Zahlen fehlen auch hier aufgrund der Geheimhaltung. Gerüchten zu Folge kann heute allerdings so gut wie jede Nachricht, die über das Internet übertragen wird grundlegend analysiert werden. Auch viele Telefonate können theoretisch mitgehört werden, da heutige Telefonverbindungen oft per angezapfter Satellitenverbindung oder Unterseekabel vermittelt werden, allerdings fällt die maschinelle, breit gefächerte Analyse und Aufzeichnung hier schwerer, da die Spracherkennung noch nicht weit genug fortgeschritten ist. Allerdings ist es mit heutiger Technik möglich Gespräche nach Stichworten auszuwerten und aufgrund dieser zur Weiterverarbeitung zu geben. Überwachungsstationen finden sich auf der ganzen Welt. Betreiber sind oft die nationalen Geheimdienste, die aber im Verbund das Gesamtsystem speisen. Diese Konstellation lässt zu, dass auch Staaten in denen man die eigenen Bürger eigentlich nicht überwachen darf, an Informationen aus ihrem Landesinneren kommen können. Überwacht werden gerüchteweise nahezu alle Unterseekabel, und ca. 120 Kommunikationssatelliten (Stand 1999). Die in Deutschland bekannteste an das System angeschlossene Station steht in Bad Aibling (Codename F-81). Die Kommunikation unter den einzelnen Abhör- und Auswertestationen erfolgt über ein dem Internet sehr ähnliches, auf TCP/IP basiertes System, welches vom öffentlichen Internet aber logischerweise vollständig abgetrennt ist.

### Ziele <sup>20 21 22 23</sup>

Über die Ziele des Programms kann nur spekuliert werden. Seit Ende des Kalten Krieges kann es nicht mehr das Hauptziel sein die Kommunikation der Sowjetunion und Chinas auszuwerten. Neuerdings ist das offizielle Hauptziel des Programmes die Bekämpfung des

---

18 Vgl. 5) Duncan Campbell (2000): „Inside Echelon“

19 Vgl. 6) Heise (2007): „Neue Hintertüren für US-Geheimdienst bei US-Telcos aufgedeckt“

20 Vgl. 2) Wikipedia (2007): „ECHELON“

21 Vgl. 6) Heise (2007): „Neue Hintertüren für US-Geheimdienst bei US-Telcos aufgedeckt“

22 Vgl. 7) Florian Rötzer (2002): „Weltweites Schnüffelsystem“

23 Vgl. 11) unbekannt (): „Echelon“

internationalen Terrorismus<sup>24</sup>, sowie die Verfolgung von Straftätern und die Aufklärung grenzüberschreitender Kriminalität. Aber auch die Überwachung von Organisationen spielt eine nicht unbedeutende Rolle. So zählen zu den Spionagezielen beispielsweise Amnesty International, Greenpeace oder religiöse Organisationen<sup>25 26</sup>. Nicht zuletzt, weil diese auch in Krisenregionen agieren, sowie Zugang zu weiteren Organisationen oder gar zu Regierungen haben. Auch die Wirtschaftsspionage kann zu den Zielen von Echelon gezählt werden. So werden zum Beispiel Daten von ausländischen Unternehmen schon mal an US-Unternehmen weitergegeben. Hier geht es um das Schaffen von Wettbewerbsvorteilen, oder das Aufholen von technologischen Lücken im eigenen Land. Aber auch als Köder, zum Gewinnen strategischer Partnerschaften, können die gewonnenen Informationen verwendet werden. Zum Beispiel wurden an die Firma Boeing im Jahre 1994 Informationen weitergegeben, die diese direkt zum Absatz eines Großauftrages nach Saudi Arabien befähigte, oder an die Rüstungsfirma Raytheon, die aufgrund von Informationen aus Echelon-Daten Radarsysteme nach Brasilien verkaufen konnte. Auch bei den Anklagen von General Motors im Verlauf der Lopez-VW Affäre sollen Daten aus dem System dem US-Konzern entscheidende Informationen geliefert haben. Ein Beispiel für Technologiespionage auf deutschem Boden lieferte der Fall Enercon<sup>27 28 29</sup>, der im Jahre 1998 von der ARD aufgedeckt wurde. So wurden dem deutschen Unternehmen Enercon per Echelon Sicherheitscodes entwendet, die von einem Agententeam verwendet wurden, um die Baupläne einer neuartigen Windkraftanlage zu entwenden. Diese wurden dem US-Unternehmen Kenetech zur Verfügung gestellt, welches die Anlage zum Patent anmeldete. Dem deutschen Unternehmen entstand ein wirtschaftlicher Verlust von ca. 100 Millionen DM.

Grenzen<sup>30 31</sup>

Grenzen des Systems lassen sich nur schwer ausmachen. Es ist erstaunlich welche Masse an Daten bereits heute überwacht wird. Für die Zukunft lässt die Verbesserung der Spracherkennung ein weiteres Wachstum der Daten vermuten. Das Nadelöhr bildet die Speicherung der Daten. Es ist nicht möglich sämtliche abgehörten Daten zu speichern, um

---

24 Vgl. 40) Institut für Geschichte, Salzburg (1999): „Big Brother is watching you“

25 Vgl. 11) unbekannt (): „Echelon“

26 Vgl. 12) Patrick S. Poole (2000): „ECHELON: America's Secret Global Surveillance Network“

27 Vgl. 10) unbekannt (): „NSA der große Bruder hört mit“

28 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

29 Vgl. 21) unbekannt (): „ECHELON“

30 Vgl. 13) Christiane Schulzki-Haddouti, Armin Medosch (1999): „Abhören im Jahr 2000“

31 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

sie dann später auszuwerten. Es muss bereits frühzeitig eine sinnvolle Filterung und Sortierung erfolgen. Dies ist umso herausfordernder, wenn man berücksichtigt, dass viele Informationen erst durch Kreuzvergleiche sichtbar werden. Es müssen also Vorgehensweisen und Taktiken erarbeitet werden, welche Daten relevant sind, und wie diese besonders günstig abgespeichert werden können, sodass vernünftige Schlüsse daraus gezogen werden können.

Interessantes <sup>32 33</sup>

Wichtig noch zu erwähnen: Die NSA wurde am 04.11.1952 auf eine Direktive des damaligen US-Präsidenten Harry S. Truman gegründet. Aufgrund dieser Besonderheit unterliegt die Institution keiner demokratischen Kontrolle. Dies lässt ungeahnte Möglichkeiten hoffen, ruft aber auch Verschwörungstheoretiker auf den Plan. Das jährliche Budget der NSA wird auf 10-20 Mrd. US-Dollar geschätzt. Genaue Zahlen sind leider nicht bekannt.

### 3. Datensammlung

Welche Daten werden gesammelt?

Laut Robert Steele, einem ehemaligen „Hacker“ der CIA, stammen ca. 40% der gesammelten Daten aus öffentlich zugänglichen Quellen wie dem Internet, aber auch aus Büchern, Lexika, Enzyklopädien und Tageszeitungen. Die restlichen 60% werden aus so bezeichneten verdeckten Quellen gewonnen <sup>34 35</sup>. Vornehmlich dienen hierzu wohl die weltweiten elektronischen Kommunikationsnetze. Abgehört werden Telefonate, Faxe, Emails, sowie sonstiger Internetverkehr. Schätzungsweise werden von den Echelon Systemen täglich drei Milliarden Nachrichten ausgewertet - Tendenz mit leistungsfähiger Hardware natürlich steigend. Ergänzt werden die automatisch gewonnenen Daten von Protokollen und Berichten, die von den Geheimdiensten erstellt und ins System gepflegt werden. Auch Trafficanalyse <sup>36</sup> spielt eine nicht unerhebliche Rolle. Aus den Bewegungsdaten und -mustern bekannter großer Netze und Netzabschnitte lassen sich Details über Informationsverbreitungen erfahren, sowie die Wege spezieller Nachrichten verfolgen. Außerdem ist es möglich Quellen und Ziele brisanter Informationen auszumachen

---

32 Vgl. 9) Ingo Ruhmann, Christiane Schulzki-Haddouti (1998): „Lauschangriff / Abhör-Dschungel“

33 Vgl. 10) unbekannt (): „NSA der große Bruder hört mit“

34 Vgl. 10) unbekannt (): „NSA der große Bruder hört mit“

35 Vgl. 11) unbekannt (): „Echelon“

36 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

und zu bewerten. Auf der Basis dieser Trafficanalyse können schließlich auch Inhalte von Nachrichten genauer beurteilt werden.

In nicht all zu ferner Zukunft wird es auch Möglichkeiten geben Fahrzeugbewegungen<sup>37</sup> in die Datenanalyse mit einzubeziehen. Hierzu werden Verkehrsüberwachungssysteme angebunden, wie jenes, das im Großraum London bereits jedes Fahrzeug erfasst, oder das deutsche Mautsystem Toll Collect, dem diese Fähigkeit vom Hersteller bestätigt wird. Ferner könnte die Gesichtserkennung<sup>38</sup> in Zukunft eine Rolle spielen. Bereits heute existieren Systeme, die einzelne Personen anhand ihrer Gesichtszüge in einer Menschenmasse identifizieren können.

Woher kommen die Daten? <sup>39 40</sup>

Bei dieser Fragestellung beschränken wir uns auf die elektronischen Abhöreinrichtungen des Systems. Das Rückgrad des Systems bildet die Überwachung der weltweiten Satelitenkommunikation. Überwacht werden primär die 120 Intelsat-Sateliten und die Inmarsat-Sateliten. Jedem dieser Systeme werden jeweils fünf Lauschposten gewidmet, die sich über den ganzen Erdball verteilen. Für die Überwachung der Intelsats sind die Großstationen in Morwenstow (GB), Sugar Grove (West Virginia/US), Yakima (Washington/US), Waihopa (Neu Seeland) und Geraldton (Australien) zuständig. Die Inmarsats werden belauscht von den Posten in Bad Aibling (Deutschland), Menwith Hill (GB), Shoal Bay (Australien), Leitrim (Kanada) und Miasawa (Japan).

Weitere Daten werden aus der Überwachung von Kabelverbindungen gewonnen. Besondere Bedeutung haben hier Unterseekabel – im Falle der Transatlantikkabel <sup>41 42 43</sup> lediglich zehn an der Zahl. Heute ist bekannt, dass zu Zeiten des Kalten Kriegs Langzeitrekorder von U-Booten an Unterseekabeln angebracht wurden. Heute dürfte dieses Vorgehen obsolet sein, zumal auch das Datenvolumen mit der Verwendung von Glasfaserkabeln deutlich angestiegen ist. Vermutlich werden Daten direkt an den Anschlussstellen abgegriffen und zur Weiterverarbeitung ins System gespeist. Relativ leicht kommt man auch an Nachrichten, die per Mobil- oder Richtfunk <sup>44 45</sup> übertragen werden. Elektromagnetische Strahlung lässt

---

37 Vgl. 40) Institut für Geschichte, Slazburg (1999): „Big Broter ist watching you“

38 Vgl. 40) Institut für Geschichte, Slazburg (1999): „Big Broter ist watching you“

39 Vgl. 11) unbekannt (): „Echelon“

40 Vgl. 40) Institut für Geschichte, Slazburg (1999): „Big Broter ist watching you“

41 Vgl. 19) unbekant (): „wie arbeitet echelon?“

42 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

43 Vgl. 5) Duncan Campbell (2000): „Inside Echelon“

44 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

45 Vgl. 9) Ingo Ruhmann, Christiane Schulzki-Haddouti (1998): „Lauschangriff / Abhör-Dschungel“

sich relativ leicht auffangen, gerüchtweise strahlen einige Richtfunkverbindungen sogar so stark ab, dass sich die Nutzlast noch im Weltraum abfangen lässt. Besonders brisant ist das Abhören des weltweiten Internet<sup>46 47 48 49</sup>. Die Topologie des Internet gerät immer wieder auf Grund ihres anfälligen Aufbaus in die Kritik. So lassen sich Flaschenhälse ausmachen und gezielt ausspionieren. Zum Beispiel wird ca. 80% des Deutschen Internetverkehrs über den DE-CIX Knoten in Frankfurt geroutet. Bereits 1998 wurde bekannt, dass die NSA ganz in der Nähe der damaligen Hauptpost in Frankfurt ein Büro mit sehr viel Technik unterhielt. Als offizieller Mieter der Räumlichkeiten entpuppte sich übrigens der deutsche BND.

#### Art der Daten<sup>50 51</sup>

Wie im Kapitel 5 (Software) noch einmal verdeutlicht wird, liegen die meisten erfassten Daten als Volltextdaten vor. Von den schätzungsweise 3 Mrd. Nachrichten pro Tag schaffen es nur sehr wenige durch die automatischen Filter bis zu einer dauerhaften Speicherung. Im Falle einer Weiterverwertung werden Nachrichten indexiert und verschlagwortet. Die genauen Vorgänge der Datenaufbereitung inklusive Dechiffrierung werden in Kapitel 4 behandelt. Über die Menge der gesammelten Daten lässt sich nur schwer etwas sagen. Laut Informationen aus dem Jahre 1998 umfasste das zentrale Datenspeichersystem der NSA zu der Zeit ca. vier Terrabyte. Heute, zehn Jahre später, werden die Speicherkapazitäten sicherlich deutlich höher sein. Nicht zuletzt auf Grund der Möglichkeit der dezentralen Speicherung. Auch hierauf wird im Kapitel 5 noch einmal näher eingegangen.

#### 4. Datenaufbereitung<sup>52</sup>

##### Filterung<sup>53 54 55</sup>

Abgefangene Daten durchlaufen mehrere Filterprozesse. Erste Filterungen erfolgen noch vor der eigentlichen Datenaufbereitung, solange Nachrichten noch in ihrer ursprünglichen Form vorliegen. Sprachnachrichten können zum Beispiel noch vor der aufwendigen Spracherkennung nach Schlüsselwörtern gescannt werden. Außerdem lassen sich

---

46 Vgl. 2) Wikipedia (2007): „ECHELON“

47 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

48 Vgl. 5) Duncan Campbell (2000): „Inside Echelon“

49 Vgl. 6) Heise (2007): „Neue Hintertüren für US-Geheimdienst bei US-Telcos aufgedeckt“

50 Vgl. 11) unbekannt (): „Echelon“

51 Vgl. 19) unbekannt (): „wie arbeitet echelon?“

52 Vgl. 1) Wikipedia (2007): „Data Mining“

53 Vgl. 2) Wikipedia (2007): „ECHELON“

54 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

55 Vgl. 39) Memex Technology Ltd (): Memex Whitepapers

Dokumente auf Schlüsselwörter in verschiedenen Sprachen untersuchen, bevor sie überhaupt in die Auswahl kommen aufwendig übersetzt zu werden. Gleiches ist bei Fax- oder Handschriftlichen Dokumenten denkbar.

Weitere Filterprozesse werden auf die schon aufbereiteten Dokumente angewendet, wenn sie elektronisch aufbereitet und übersetzt wurden. Hierbei helfen Data Mining Methoden, wie Natural Language Processing, Text Mining, Pattern Matching, auf die im Kapitel 6 näher eingegangen werden soll. Auch KI (künstliche Intelligenz) findet bei der Filterung der Nachrichten weitflächige Anwendung. Die Ergebnisse der Filterung werden von Operatoren bewertet und somit ein Lernprozess innerhalb der KI angestoßen. Künstliche neuronale Netze eignen sich hierzu in besonderem Maße. Zu den trainierenden Prozessen gehört hier nicht nur die manuelle Bewertung der Ergebnisse, sondern auch die automatische Analyse von Workflows, sowie der Workflowdurchsatz. So können aufgrund der Art der Daten und der Geschwindigkeit von Workflowdurchläufen Rückschlüsse auf die Wichtigkeit von Rohdaten gezogen werden. Diese Ergebnisse können schließlich erneut zum Training der Filterprozesse herangezogen werden. Eine in Anbetracht der sonstigen Komplexität eher banal anmutende, aber keinesfalls unwichtige Filteraufgabe ist das Eliminieren doppelter Einträge. Auch das Kennzeichnen von Nachrichten anhand eines Schlüsselwortkataloges<sup>56</sup> zählt eher zu den Routineaufgaben. Für die Weiterverarbeitung der Daten jedoch ist dieser Schritt von enormer Wichtigkeit.

#### Verteilung der Daten<sup>57 58</sup>

Die gewonnenen rudimentär aufbereiteten Daten werden je nach Interessensgebiet automatisch an die weiterverarbeitenden Institutionen (Agencies) weitergeleitet. Jede Institution unterhält Schlüsselwortkataloge, anhand derer die für sie wichtigen Daten aus den Rohdaten gewonnen werden. Nach dem Tagging (Kennzeichnen) werden diese für den dezentralen Zugriff über ein „Intelink“ genanntes System vorbereitet. In einem weiteren Verarbeitungsschritt, wird dann händisch die Bewertung der Nachrichten durchgeführt, um die KI zu speisen. Die Verfolgung der Workflows geschieht automatisch.

---

56 Vgl. 21) unbekannt (): „ECHELON“

57 Vgl. 40) Institut für Geschichte, Slazburg (1999): „Big Broter ist watching you“

58 Vgl. 27) Detlef Borchers (1999): „Agenten im Netz“

OCR, Spracherkennung, Übersetzung<sup>59 60 61 62</sup>

Um Faxnachrichten elektronisch auswertbar zu machen kommen OCR-Verfahren (Optical Character Recognition) zum Einsatz. Diese sind bereits in zivilen Versionen sehr ausgereift, eine Herausforderung dürfte lediglich die Anwendung auf riesige Datenmassen sein, sowie die Portierung der Verfahren auf Großrechner und deren Optimale Auslastung. Dies ist aber nur am Rande ein Problem des Beschäftigungsfeldes Data Mining. Anspruchsvoller sind Methoden, wie die Spracherkennung, die auf Telefonate angewandt werden, ebenso wie die korrekte Übersetzung aus anderen Sprachen. Die beiden letztgenannten Problematiken gehen Hand in Hand. Theoretisch ist es, leistungsfähige Hardware vorausgesetzt, nicht sonderlich schwer gesprochene Sprache in elektronisch verwertbaren Volltext umzusetzen (Natural Language Processing), allerdings steigt der Aufwand, je mehr Sprachen und darin enthaltene Wörter erkannt werden sollen. Auch die Tatsache, dass sprecherunabhängig Sprache erkannt werden muss erhöht den Aufwand. Ähnlich komplex sind die Bemühungen um vernünftige Übersetzungslösungen. Reine wörtliche Übersetzungen sind für den Menschen kaum zu gebrauchen, maschinell lassen sich diese Übersetzungen jedoch einigermaßen brauchbar verwenden. Sowohl beim Natural Language Processing als auch zur Spracherkennung werden verborgene Markow Modelle und neuronale Netze eingesetzt. Näheres dazu lesen Sie in Kapitel 6 (Data Mining Techniken).

Entschlüsselung<sup>63 64 65 66 67 68</sup>

Fakten zur Nachrichtenentschlüsselung im Zusammenhang mit Echelon lassen sich nur wenige finden. Bekannt ist, dass die NSA einer der weltgrößten Arbeitgeber für Mathematiker ist, die in Spezialabteilungen als „Codebreaker“ arbeiten. Dort beschäftigt man sich mit Schwachstellen in gängigen Verschlüsselungsalgorithmen und deren gezielter Ausnutzung. Auch, dass Computertechnologie vorhanden ist, um per Bruteforce Verschlüsselungen zu knacken ist weitgehend bekannt. Als Beispiel wäre hier die EFF-Maschine (Electronic Frontier Foundation) zu nennen, die einzig und allein dafür entwickelt wurde den Verschlüsselungsalgorithmus „DES“ zu knacken. Ähnliche Technologie dürfte

---

59 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

60 Vgl. 15) unbekannt (2005): „Neuigkeiten von der ECHELON-Front“

61 Vgl. 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“

62 Vgl. 12) Patrick S. Poole (2000): „ECHELON: America's Secret Global Surveillance Network“

63 Vgl. 22) Dirk Fox (): „Editorial: Im Geschwindigkeitsrausch“

64 Vgl. 23) Besim Karadeniz (): „Das meistgesuchtete Hintertor“

65 Vgl. 24) Wikipedia (2007): „NSAKEY“

66 Vgl. 12) Patrick S. Poole (2000): „ECHELON: America's Secret Global Surveillance Network“

67 Vgl. 21) unbekannt (): „ECHELON“

68 Vgl 20) Kai Billen (): „ENFOPOL – das europäische Abhörnetzwerk“

auch bei Echelon im Einsatz sein. Da Bruteforcedechiffrierung allerdings einen sehr hohen Aufwand darstellt, darf davon Ausgegangen werden, dass nur sehr wenige verschlüsselte Nachrichten auf diese Weise geknackt werden. Populärer ist das Finden und Ausnutzen von Sicherheitslücken – nicht nur der Verschlüsselungsalgorithmen selbst. Gerüchteweise pflegt die NSA intensive Kontakte mit Herstellern von Standardsoftware. So verwundert es doch sehr, dass bekannte Groupwarelösungen, wie IBM Lotus Notes, oder Microsoft Outlook, die in vielen Firmen und Institutionen zum Versenden vertraulicher Nachrichten verwendet werden, im Auslieferungszustand keine Verschlüsselungsmechanismen mitliefern. Auch wurde im Jahr 1999 vom Wissenschaftler Andrew D. Fernandez in der mit allen Windows-Systemen ausgelieferten Microsoft Crypto-API eine Hintertür aufgedeckt, die den Microsoft internen ominösen Namen „\_NSAKEY“ trägt. Eine Zusammenarbeit von Microsoft und NSA wird selbstverständlich von beiden Parteien bis heute abgestritten.

## 5. Software

### MEMEX <sup>69</sup> <sup>70</sup>

Die Firma Memex aus Großbritannien steuert dem Echelonssystem das intelligente Rastersystem bei, das für die Datenvorauswahl, die Datenaufbereitung und die Datenanalyse zuständig ist. Memex vertreibt, über ihre Internetseite ersichtlich, eine Produktreihe aus dem Bereich der künstlichen Intelligenz, die das Leistungsvermögen des an die NSA gelieferten Systems wahrscheinlich sehr genau widerspiegelt. Laut den öffentlich zugänglichen Whitepapers Handelt es sich bei der derzeit verfügbaren Produktreihe Memex Series VI um eine integrierte Softwarelösung zur Anwendung von künstlicher Intelligenz auf das Durchsuchen von Volltextdatenbanken. Weiter ist den Funktionsbeschreibungen zu entnehmen, dass das System sich um die Verschlagwortung (Keywords) und die Einordnung von Texten oder Textteilen in Kategorien kümmert. Auch wirbt man mit besonderer Flexibilität bei der Anbindung unterschiedlicher Datenquellen, was den Anforderungen von Echelon auf Grund seiner Verteilten Anordnung sehr entgegenkommen dürfte. So können laut Memex sowohl herkömmliche relationale Datenbanken, als auch Volltextdaten, sowie moderne Datenaustauschformate wie XML angebunden werden. Memex selbst ist Oracle Partner und bezeichnet sein Produkt als Datenbankhybrid. Es stellt eine virtuelle zentralisierte Datenbank bereit. Was wiederum genau dem Aufbau von Echelon entspricht.

---

69 Vgl. 3) Kai Billen (): „Echelon – Das globale Abhörnetzwerk“ (Seite 2 - Memex)

70 Vgl. 39) Memex Technology Ltd (): Memex Whitepapers

Weiter lassen sich an Memex Series VI auch externe Tools anbinden. Als Beispiel werden GIS-Tools genannt, oder von Memex selbst gelieferte Zusatzprogramme zur grafischen Auswertung der Analysedaten. Auf Grund der vielen Schnittstellen ist aber auch die Anbindung von Analysetools anderer Hersteller denkbar. Schätzungsweise wird auch das Intranet der US-Geheimdienste, das Intelink, mit Memex-Daten gespeist.

Memex Series VI besteht aus drei Kernprodukten. Das erste ist Memex Patriarch. Es ist zuständig für die Serverseitige Datenhaltung, Datenaufbereitung und Rechteverwaltung. Außerdem bietet es ein Workflowmanagement. Wie bereits im Kapitel 4 erwähnt fließen die Daten aus der Workflowkontrolle ebenfalls in das intelligente Rastersystem ein. Das zweite Teilprogramm ist Memex Sententia. Sententia ist zuständig für die webbasierte Präsentation der Daten und ermöglicht so einen verteilten Zugriff auf alle Geheimdienstdaten. Das dritte und vermutlich wichtigste Programm ist Memex Analyst. Es ermöglicht Datenanalyse in Echtzeit und bietet Werkzeuge zur Entscheidungshilfe und Entscheidungsfindung mit Hilfe künstlicher Intelligenz und Neuronaler Netze. Hierzu lesen Sie mehr im nächsten Kapitel.

Intelink <sup>71 72 73 74 75</sup>

Intelink ist das Intranet der Geheimdienste, das seit dem Jahre 1996 existiert. Seit Anfang der neunziger Jahre gibt es bei den amerikanischen Geheimdiensten die Bestrebung eine einheitliche Dateninformationsbasis zu schaffen. Zuerst wurde versucht dies mit Hilfe von Comuserve angekaufter Hardware und Know How zu bewerkstelligen. Hierbei scheiterte man allerdings an der Heterogenität der Ausrüstung der unterschiedlichen Abteilungen. Erst die technische Basis, die auch im Internet verwendet wird – TCP/IP & Hypertext Markup Language – brachte den entscheidenden Durchbruch. Nicht nur technisch, sondern auch finanziell. So werden, wie bei der Verleihung des Hammer Awards bekannt wurde, jährlich etwa 5 Mrd. Dollar durch die Verwendung des Systems eingespart. Besonderes Augenmerk aus technischer Sicht gilt natürlich der Sicherheit des Netzes selbst, sowie der Rechteverwaltung innerhalb des Systems. So ist Intelink in mehrere Sicherheits- und Vertraulichkeitsstufen unterteilt – von öffentlich zugänglich bis „top secret“.

Die Vorteile liegen klar auf der Hand; einerseits ermöglicht ein Intranet die Veröffentlichung und Verbreitung von Informationen „just in time“, andererseits ist es ohne große

---

71 Vgl. 25) Wikipedia (2007): „Intelink“

72 Vgl. 26) John Pike (2003): „Intelink“

73 Vgl. 27) Detlef Borchers (1999): „Agenten im Netz“

74 Vgl. 28) Emmett Paige Jr. (1996): „The rapid expansion of Intelink“

75 Vgl. 29) unbekannt (): „Directory of Intelink Servers“

Veränderungen der Kerntechnologie nach oben skalierbar. Ersteres dürfte besonders im operativen Einsatz des Militärs und der Geheimdienste entscheidende strategische Vorteile bieten. Das Datenvolumen des Netzes betrug im Jahre 1999 bereits mehrere Petabytes ( $10^{15}$  Bytes), und dürfte auf Grund der einfachen Skalierbarkeit und den Bemühungen immer mehr Institutionen anzubinden heute wesentlich höher sein. Für die einheitliche Darstellung werden offene Datenstandards genutzt. Für strukturierte Textinformationen beispielsweise SGML (Standard Generalized Markup Language). Hierbei darf vermutet werden, dass Intelink und Memex diese Formate zum Datenaustausch verwenden.

Zum Durchsuchen der riesigen Datenmengen werden sowohl kommerzielle Programme verwendet, als auch selbst entwickelte Suchmaschinen, die den Anforderungen der Geheimdienste genügen. Zur Informationsaufbereitung werden Methoden wie Clustering verwandt, sowie die Anwendung von Metadaten.

Direkten Nutzen aus Intelink ziehen z.B. Operativ agierende Einheiten mit dem FALCON (Forward Area Language Converter). Diese Laptop-Scanner-Kombination ermöglicht es ein in beliebiger Sprache vorliegendes Dokument einzulesen und dem Anwender Informationen über den Inhalt zu liefern. Auch wenn das ursprüngliche Dokument in einer dem Anwender völlig unbekanntem Sprache wie Chinesisch oder Kyrillisch geschrieben war.

Intelink selbst setzt außer bei der Datenaufbereitung auch bei den verwendeten Programmen auf Standards. So werden z.B. Browser von Netscape oder Microsoft verwendet, und Sprachkommunikation findet per VOIP statt. Unter anderem deshalb ist die immense Kosteneinsparung überhaupt nur möglich.

## 6. Data Mining Techniken<sup>76</sup>

In diesem Kapitel sollen beispielhaft einige in Echelon verwendete Techniken und Methoden des Data Mining vorgestellt werden. Insbesondere werden hierzu Methoden gewählt, die in den vorherigen Kapiteln bereits erwähnt wurden, oder diese erklären und ergänzen können. Die Ausführung erhebt keinen Anspruch auf Vollständigkeit, dies würde den Rahmen dieser Arbeit weit überschreiten.

### 1. Indexierung<sup>77 78</sup>

---

76 Vgl. 1) Wikipedia (2007): „Data Mining“

77 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

78 Vgl. 30) Wikipedia (2007): „Indexierung“

„Als Indexierung oder auch Verschlagwortung (Österreich: Beschlagwortung) bezeichnet man beim Information-Retrieval die Zuordnung von Deskriptoren zu einem Dokument zur Erschließung der darin enthaltenen Sachverhalte. Es lassen sich die kontrollierte Indexierung mit einem Thesaurus oder Schlagwortkatalog bzw. Notationen einer Klassifikation und freie Indexierung bzw. freie Verschlagwortung mit nicht vorgegebenen Deskriptoren unterscheiden. Beim Gemeinschaftlichen Indexieren (auch social oder collaborative tagging) mit Hilfe von Sozialer Software spricht man auch von Tagging anstelle von Indexierung und von Tags anstatt von Deskriptoren.“

Quelle: 30) Wikipedia (2007): „Indexierung“

Bei der Indexierung wird unterschieden zwischen der manuellen Indexierung, der automatischen Indexierung und der computergestützten Indexierung. Die letztgenannte Variante stellt einen Hybrid aus den beiden anderen dar. Schlüsselwörter werden automatisch generiert, jedoch manuell validiert und bewertet.

Die Indexierung von Dokumenten dient der späteren Recherchierbarkeit des Dokuments innerhalb einer größeren Ansammlung von Dokumenten. Hierzu werden oft auch Dokumentenklassen gebildet, die nach Stichwörtern „geclustert“ werden.

So wird ein effektives Dokumentenmanagement ermöglicht. Dokumente können so sinnvoll elektronisch archiviert werden. Im betrieblichen Umfeld wird dieses Vorgehen Enterprise Content Management genannt.

Bei Echelon dient die Indexierung genau dem selben Zweck. Dokumente und Nachrichten, die über Abhöreinrichtungen oder -maßnahmen gewonnen wurden, werden für spätere Recherchearbeiten mit Hilfe von Indexierungsmaßnahmen aufbereitet. Außerdem liefert die Indexierung Metadaten, anhand welcher Mustererkennungen durchgeführt und Zusammenhänge analysiert werden können.

## 2. Metadaten <sup>79</sup>

„Als Metadaten oder Metainformationen bezeichnet man allgemein Daten, die Informationen über andere Daten enthalten. Bei den beschriebenen Daten handelt es sich oft um größere Datensammlungen (Dokumente) wie Bücher, Datenbanken oder Dateien. So werden auch Angaben von Eigenschaften eines Objektes (beispielsweise

---

<sup>79</sup> Vgl. 31) Wikipedia (2007): „Metadaten“

Personennamen) als Metadaten bezeichnet. Während der Begriff „Metadaten“ relativ neu ist, ist sein Prinzip unter anderem jahrhundertlang bibliothekarische Praxis.“

Quelle: 31) Wikipedia (2007): „Metadaten“

Metadaten dienen ähnlichen Zwecken wie die Ergebnisse der Indexierung. Oft werden Indexierungsdaten auch als Metadaten abgespeichert. Allerdings können Metadaten auch Informationen beinhalten, die nicht direkt aus dem Dokument stammen, dem sie angehängt wurden. Hierunter fallen Daten, wie die Quelle des Dokumentes, Verbreitungsinformationen oder Dringlichkeitsbewertungen.

Bei Echelon werden Metainformationen verwendet, um abgefangenen Nachrichten Informationen beizufügen, die in die weitere Analyse mit einfließen sollen. Dies können sowohl durch Menschen angefügte Daten, wie Bewertungen oder Berichte sein, als auch maschinell hinzugefügte Daten, wie die automatisch ermittelte Quelle der Nachricht, das Abfangdatum oder zugeordnete Schlüsselwörter aus abteilungsspezifischen Wörterbüchern.

### 3. Text Mining <sup>80</sup>

„Der Begriff Textmining (engl. text mining, von text data mining) bezeichnet die automatisierte Entdeckung neuer, richtiger und relevanter Informationen aus Textdaten. Mit statistischen und linguistischen Mitteln erschließt die Textmining-Software aus Texten Informationen, die die Benutzer in die Lage versetzen soll, ihr Wissen zu erweitern oder ihre Handlungen daran auszurichten. Textmining-Systeme liefern im Optimalfall Informationen, von denen Benutzer bisher nicht wussten, dass sie sie nicht kannten. Im Zusammenspiel mit ihren Anwendern sind Werkzeuge des Textminings außerdem dazu in der Lage, Hypothesen zu generieren, diese zu überprüfen und schrittweise zu verfeinern. Textmining zählt deshalb auch zu den Verfahren der explorativen Datenanalyse.“

Quelle: 32) Wikipedia (2007): „Textmining“

Textmining beschäftigt sich im Gegensatz zum Information Retrieval (IR) mit der Informationsgewinnung aus einem einzigen Volltext. IR hingegen erhebt den Anspruch

---

<sup>80</sup> Vgl. 32) Wikipedia (2007): „Textmining“

aus einer Masse von Dokumenten die treffendsten zu liefern. Weiter ist es ein wichtiges Ziel des Textmining besonders solche Informationen aus einem Text zu extrahieren, die dem Nutzer noch nicht bekannt waren. Besonderes Augenmerk im Zusammenhang mit Textmining kommt der linguistischen Analyse zu. So ist es äußerst wichtig mehrdeutige Wörter herauszufinden und anhand des restlichen Kontextes korrekt zu bestimmen.

Bei Echelon wird Text Mining zur Analyse abgefangener Nachrichten verwendet. Sowohl vor dem eigentlichen Filterprozess kann Text Mining wichtige Informationen zur Einstufung einer Nachricht liefern, aber auch im weiteren Analyseverlauf kann auf Text Mining zurückgegriffen werden um weiterführende Informationen aus einem Volltext zu gewinnen. Auch beim Herausfinden von Trends spielt Text Mining im weiteren Analyseverlauf eine große Rolle.

#### 4. Natural Language Processing (NLP) <sup>81 82</sup>

„Der Begriff Verarbeitung Natürlicher Sprache (VNS; engl. natural language processing, NLP, daher oftmals falsch als Natürliche Sprachverarbeitung übersetzt) bezeichnet ein Forschungsgebiet, das Einsichten der Computerlinguistik zur Spracherkennung und Sprachsynthese einsetzt. VNS wird im Deutschen oft als Sprachverstehen bezeichnet.“  
Quelle: 33) Wikipedia (2007): „Natural language processing“

Zum maschinellen Verstehen von Sprache werden häufig neuronale Netze und verborgene Markow-Modelle verwendet. Beide Techniken werden im Verlauf des Kapitels kurz vorgestellt werden. Weitere wichtige Schlüsseltechnologien des NLP sind die Mustererkennung und die Musterbildung. Auch diese werden in diesem Kapitel vorgestellt.

Im Rahmen von Echelon spielt das maschinelle Verstehen der natürlichen Sprache , wenn auch nur in naher Zukunft, eine eklatante Rolle. Will man Computersysteme dafür verwenden automatisch aus einer großen Masse von aufgezeichneten Nachrichten relevante Informationen herauszufiltern, so müssen die Computer diese Nachrichten zuerst einmal verstehen. NLP kann so weitere Metadaten für die nachfolgende Analyse

---

81 Vgl. 18) Kai Billen (): „COMINT und UKUSA-Abkommen“

82 Vgl. 33) Wikipedia (2007): „Natural language processing“

der Nachrichten liefern.

## 5. Pattern Matching<sup>83 84</sup>

„Pattern Matching (engl. für Musterabgleich) ist ein mathematischer Suchalgorithmus. Das Verfahren ermittelt in endlicher Zeit, ob sich ein gegebenes Muster (Pattern) in einem (begrenzten) Suchbereich wiederfindet. Im Gegensatz zur Mustererkennung (dem Finden von Mustern in Signalen) wird der Pattern vorher angegeben.“

Quelle: 34) Wikipedia (2007): „Pattern Matching“

Pattern Matching stellt die einfachste Form der Analyse dar, die zum Filtern abgefangener Nachrichten verwendet wird. Klartextnachrichten, oder solche, die bereits entschlüsselt sind können per Pattern Matching einfach auf Stichwörter oder Varianten von Stichwörtern durchsucht werden. Wird ein Stichwort gefunden, so kann das Dokument ohne großen weiteren Analyseaufwand direkt der Auswertung zugeführt werden.

## 6. Mustererkennung<sup>85 86</sup>

„Die Mustererkennung, ein Teilgebiet der Informatik, untersucht Verfahren, die gemessene Signale automatisch in Kategorien einordnen. Zentraler Punkt ist dabei das Erkennen von Mustern, den Merkmalen, die allen Dingen einer Kategorie gemeinsam sind und sie vom Inhalt anderer Kategorien unterscheiden. Mustererkennungsverfahren befähigen Computer, Maschinen und Roboter, statt präziser Eingaben auch die weniger exakten Eindrücke einer natürlichen Umgebung zu verarbeiten.“

Quelle: 35) Wikipedia (2007): „Mustererkennung“

Im Gegensatz zum Pattern Matching werden bei der Mustererkennung die Muster aus den vorliegenden Volltextdaten generiert. Man unterscheidet grundsätzlich zwischen der syntaktischen Mustererkennung, der statistischen, sowie der strukturellen. Die syntaktische Mustererkennung versucht im Falle eines Fließtextes diesem Folgen von

---

83 Vgl. 12) Patrick S. Poole (2000): „ECHELON: America's Secret Global Surveillance Network“

84 Vgl. 34) Wikipedia (2007): „Pattern Matching“

85 Vgl. 7) Florian Rötzer (2002): „Weltweites Schnüffelsystem“

86 Vgl. 35) Wikipedia (2007): „Mustererkennung“

Merkmale zu entnehmen und anhand dieser Merkmale Vergleiche mit anderen Texten oder Textpassagen anzustellen. Da es allerdings oft schwer ist Merkmale klar voneinander zu trennen, findet dieser Ansatz heute nur noch wenig Bedeutung. Wesentlich öfter kommt die statistische Mustererkennung zum Einsatz. Hierbei wird nicht nach Merkmalen direkt gesucht, sondern es wird anhand von kumulierten granularen Wahrscheinlichkeiten bestimmt, welcher Kategorie ein Objekt zugeordnet werden kann. Die dritte Art der Mustererkennung, die strukturelle, ist eine Zusammenführung der beiden anderen. Je nach Art eines Merkmals wird dabei die syntaktische oder die statistische Mustererkennung angewandt, die beiden verschiedenen Ergebnistypen werden schließlich mit Hilfe von Bayesschen Netzen wieder zusammengeführt. Das Vorgehen bei der Mustererkennung gliedert sich normalerweise in die Schritte der Vorverarbeitung, wobei unerwünschte Daten reduziert werden, die Merkmalsgewinnung, bei der auf Grund früher stattgefundener Lernprozesse entschieden wird, welche Merkmale des vorliegenden Textes essentiell sein könnten, die Merkmalsreduktion, bei der die zuvor gewonnenen Merkmale mit Hilfe von Varianzanalysen wieder reduziert werden, und schließlich die Klassifikation. Beim letzten und wichtigsten Schritt werden die Merkmale schlussendlich in Klassen eingeordnet. Bei Bedarf wird noch versucht das erzeugte Muster zu interpretieren, was aber Aufgabe eines eigenen Wissensgebiet, der Musteranalyse, ist.

Im Echelonssystem wird die Mustererkennung eingesetzt im Rahmen der Spracherkennung, der Texterkennung und in Zukunft vermutlich auch bei der Gesichtserkennung. Inwiefern Mustererkennungsalgorithmen im Rahmen der KI-Engine der Software Memex Series VI verwendet werden war den Produktbeschreibungen nicht klar zu entnehmen, allerdings ist die Verwendung von Mustererkennung beim Lernprozess künstlicher intelligenter Systeme quasi unumgänglich.

## 7. Verborgene Markow Modelle <sup>87 88</sup>

„Das Verborgene Markow-Modell (VMM, engl. Hidden Markov Model, HMM), benannt nach dem russischen Mathematiker Andrei Andrejewitsch Markow, ist ein stochastisches Modell, das sich durch zwei Zufallsprozesse beschreiben lässt.

---

87 Vgl. 33) Wikipedia (2007): „Natural Language Processing“

88 Vgl. 36) Wikipedia (2007): „Verborgenes Markow-Modell“

Der erste Zufallsprozess entspricht dabei einer Markow-Kette, die durch Zustände und Übergangswahrscheinlichkeiten gekennzeichnet ist. Die Zustände der Kette sind von außen jedoch nicht direkt sichtbar (sie sind verborgen, hidden). Stattdessen erzeugt ein zweiter Zufallsprozess zu jedem Zeitpunkt beobachtbare Ausgangssymbole gemäß einer zustandsabhängigen Wahrscheinlichkeitsverteilung. Die Aufgabe besteht häufig darin, aus der Sequenz der Ausgabesymbole auf die Sequenz der verborgenen Zustände zu schließen.“

Quelle: 36) Wikipedia (2007): „Verborgenes Markow-Modell“

Verborgene Markow-Modelle (VMM) werden häufig zur Mustererkennung eingesetzt, besonders wenn Daten in einem kontinuierlichen sequentiellen Datenstrom eintreffen. So werden VMM häufig zur Spracherkennung eingesetzt. Aber auch zur Schrifterkennung und zur psychologischen Analyse können VMM verwendet werden.

Im Rahmen von Echelon können VMM zur Mustererkennung beim Data Mining selbst, beim Filtern von Datenströmen, und zum Analysieren von Sprachsignalen verwendet werden. Eventuell werden sie auch zur Schrifterkennung eingesetzt.

## 8. Künstliche Intelligenz<sup>89 90 91</sup>

„Künstliche Intelligenz (KI, engl. artificial intelligence, AI) ist ein Teilgebiet der Informatik, das sich mit der Automatisierung intelligenten Verhaltens befasst. Der Begriff ist insofern schwierig, als es keine genaue Definition von Intelligenz gibt. Trotzdem findet er in der Forschung und Entwicklung Anwendung.“

Quelle: 37) Wikipedia (2007): „Künstliche Intelligenz“

Das Ziel Künstlicher Intelligenz (KI) ist oft die Nachahmung des Menschen. Auf Grund der Tatsache, dass es äußerst schwer ist den Menschen überhaupt zu verstehen, stellt sich dieses Forschungsgebiet eher als visionäre, denn als praktische Variante heraus. Das Nachempfinden menschlicher Intelligenz wird als starke KI bezeichnet. Im Gegensatz dazu kümmert sich eine schwache KI lediglich um konkrete Anwendungsprobleme. Hier genügt es wichtige Intelligenzmerkmale zu isolieren und gegebenenfalls zu abstrahieren.

---

89 Vgl. 9) Ingo Ruhmann, Christiane Schulzki-Haddouti (1998): „Lauschangriff / Abhör-Dschungel“

90 Vgl. 39) Memex Technology Ltd (): Memex Whitepapers

91 Vgl. 37) Wikipedia (2007): „Künstliche Intelligenz“

Die Forschung und Entwicklung schwacher KI's stellt sich als wesentlich erfolgversprechender heraus. Zur Realisierung künstlicher Intelligenz werden häufig künstliche neuronale Netze verwendet.

Künstliche Intelligenz wird als eine der Kerneigenschaften des Intelligenten Rastersystems Memex Series VI genannt. Es ist davon auszugehen, dass im Rahmen der Nutzung von Memex-Software im System Echelon auch Techniken aus dem Bereich der künstlichen Intelligenz zum Einsatz kommen.

## 9. Künstliche Neuronale Netze <sup>92 93</sup>

„Künstliche neuronale Netze (kurz: KNN, engl. artificial neural network – ANN) sind Netze aus künstlichen Neuronen. Sie sind ein Zweig der künstlichen Intelligenz und prinzipieller Forschungsgegenstand der Neuroinformatik. Der Ursprung der künstlichen neuronalen Netze liegt ebenso, wie bei den künstlichen Neuronen, in der Biologie. Man stellt sie den natürlichen neuronalen Netzen gegenüber, welche Nervenzellvernetzungen im Gehirn und im Rückenmark bilden. Insgesamt geht es aber um eine Abstraktion (Modellbildung) von Informationsverarbeitung und weniger um das Nachbilden biologischer neuronaler Netze.“

Quelle: 38) Wikipedia (2007): „Künstliche neuronale Netze“

Ein wesentliches Merkmal von künstlichen neuronalen Netzen (KNN) ist ihre Lernfähigkeit. Ein Lernprozess trainiert ein KNN darauf bestimmten Eingaben passende Ausgaben zuzuordnen. Dieses Lernen kann überwacht, oder unüberwacht statt finden. Beim überwachten Lernen werden vom KNN produzierte Ausgaben mit Sollwerten verglichen. Durch Vergleich der beiden Größen kann die am KNN vorzunehmende Änderung bestimmt werden. Beim unüberwachten Lernen werden lediglich Eingabewerte geliefert, das Netz strukturiert sich autonom neu. Eine Zwischenform der beiden vorigen Methoden ist das bestärkende Lernen. Hierbei handelt das KNN zunächst selbständig, allerdings werden nur Teilschritte vollzogen. Diese Teilschritte werden progressiv bewertet, wodurch Rückschlüsse auf die noch verbleibenden Teilschritte gezogen werden können. KNN können eingesetzt werden zur Regelung von

---

92 Vgl. 33) Wikipedia (2007): „Natural Language Processing“

93 Vgl. 38) Wikipedia (2007): „Künstliche neuronale Netze“

komplexen Prozessen, zur Realisierung von Frühwarnsystemen, zur Optimierung komplexer Sachverhalte, sowie zur Mustererkennung, zum Beispiel bei der Schrifterkennung (OCR), der Spracherkennung, und allgemein beim Data Mining.

Echelon bedient sich KNN vermutlich bei der Schrifterkennung und der Spracherkennung. Möglicherweise werden KNN aber auch zur Datenanalyse, der Optimierung der Informationsaufbereitung und für angeschlossene Frühwarnsysteme verwendet.

## 7. Schlussfolgerung

Über Echelon gibt es viele Legenden und leider wenige fundierte Kenntnisse. Auf viele dieser Legenden geht diese Arbeit ein, hat jedoch nicht die Absicht diese zu bestätigen. Es soll lediglich aufgezeigt werden, dass viele Mutmaßungen mit den richtigen Mitteln bereits heute leicht realisierbar sind. Ob dies so von den Betreibern von Echelon tatsächlich umgesetzt wird bleibt weiter dahingestellt.

## 8. Quellenangaben

- 1) Wikipedia (Dez. 2007): „Data Mining“: <http://de.wikipedia.org/wiki/Data-Mining> (12.2007)
- 2) Wikipedia (Dez. 2007): „ECHELON“: <http://de.wikipedia.org/wiki/ECHELON> (12.2007)
- 3) Kai Billen (): „Echelon – Das globale Abhörnetzwerk“:  
<http://hp.kairaven.de/miniwahr/echelon-index.html> (12.2007)
- 4) Reinhard Wobst (): „Über technische Möglichkeiten und Grenzen großer Geheimdienste“:  
<http://squat.net/gib/echelon/technisches.html> (12.2007)
- 5) Duncan Campbell (2000): „Inside Echelon“:  
<http://www.heise.de/tp/r4/artikel/6/6928/1.html> (12.2007)
- 6) Heise (2007): „Neue Hintertüren für US-Geheimdienst bei US-Telcos aufgedeckt“: <http://www.heise.de/newsticker/meldung/100662/from/rss09> (12.2007)
- 7) Florian Rötzer (2002): „Weltweites Schnüffelsystem“:  
<http://www.heise.de/tp/r4/artikel/13/13647/1.html> (12.2007)
- 8) Nicky Hager (2000): „Wie ich Echelon erforscht habe“: <http://www.heise.de/tp/r4/artikel/>

[6/6728/1.html](#) (12.2007)

9) Ingo Ruhmann, Christiane Schulzki-Haddouti (1998): „Lauschangriff / Abhör-Dschungel“: <http://www.heise.de/ct/98/05/082/> (12.2007)

10) unbekannt (): „NSA der große Bruder hört mit“:

[http://www-pu.informatik.uni-tuebingen.de/iug/archiv/SoSe00/sose00\\_NSA.htm](http://www-pu.informatik.uni-tuebingen.de/iug/archiv/SoSe00/sose00_NSA.htm) (12.2007)

11) unbekannt (): „Echelon“: <http://www.verschwoerungen.info/wiki/Echelon> (12.2007)

12) Patrick S. Poole (2000): „ECHELON: America's Secret Global Surveillance Network“: <http://fly.hiwaay.net/%7Epspoole/echelon.html> (12.2007)

13) Christiane Schulzki-Haddouti, Armin Medosch (1999): „Abhören im Jahr 2000“: <http://www.heise.de/tp/r4/artikel/2/2833/1.html> (12.2007)

14) Kai Billen (): „Echelon und die europäische Union“:

<http://hp.kairaven.de/miniwahr/echelon4.html> (12.2007)

15) unbekannt (2005): „Neuigkeiten von der ECHELON-Front“: <http://confidenz-depesche.com/cdpub/cdabo9907/cd9907bb02.html> (12.2007)

16) Wikipedia (2007): „UKUSA“: <http://de.wikipedia.org/wiki/UKUSA> (12.2007)

17) Philippe Riviere (1999): „Echelon – Globale Überwachung unter der Regie der USA“: <http://www.trend.infopartisan.net/trd0199/t340199.html> (12.2007)

18) Kai Billen (): „COMINT und UKUSA-Abkommen“:

<http://hp.kairaven.de/miniwahr/echelon2.html> (12.2007)

19) unbekannt (): „wie arbeitet echelon?“:

[http://www.hackersunderground.net/themen/bigbrother/echelon/2\\_wie.htm](http://www.hackersunderground.net/themen/bigbrother/echelon/2_wie.htm) (12.2007)

20) Kai Billen (): „ENFOPOL - das europäische Abhörnetzwerk“:

<http://hp.kairaven.de/miniwahr/enfopol.html> (12.2007)

21) unbekannt (): „ECHELON“: <http://staatsfeind.net/ECHELON/echelon.html> (12.2007)

22) Dirk Fox (): „Editorial: Im Geschwindigkeitsrausch“: <http://datenschutz-und-datensicherheit.de/jhrg23/edit9910.htm> (12.2007)

23) Besim Karadeniz (): „Das meistgesuchtete Hintertor“: <http://www.netplanet.org/i-files/file012.shtml> (12.2007)

24) Wikipedia (2007): „\_NSAKEY“: <http://en.wikipedia.org/wiki/NSAKEY> (12.2007)

25) Wikipedia (2007): „Intelink“: <http://en.wikipedia.org/wiki/Intelink> (12.2007)

26) John Pike (2003): „Intelink“: <http://www.fas.org/irp/program/disseminate/intelink.htm> (12.2007)

27) Detlef Borchers (1999): „Agenten im Netz“:

[http://images.zeit.de/text/1999/08/199908.comp\\_intelink\\_.xml](http://images.zeit.de/text/1999/08/199908.comp_intelink_.xml) (12.2007)

- 28) Emmett Paige Jr. (1996): „The rapid expansion of Intelink“:  
<http://www.fas.org/irp/program/disseminate/di1166.htm> (12.2007)
- 29) unbekannt (): „Directory of Intelink Servers“:  
<http://www.fas.org/irp/program/disseminate/docs/intelink-list.htm> (12.2007)
- 30) Wikipedia (2007): „Indexierung“: <http://de.wikipedia.org/wiki/Indexierung> (12.2007)
- 31) Wikipedia (2007): „Metadaten“: <http://de.wikipedia.org/wiki/Metadaten> (12.2007)
- 32) Wikipedia (2007): „Textmining“: <http://de.wikipedia.org/wiki/Textmining> (12.2007)
- 33) Wikipedia (2007): „Natural language processing“:  
[http://de.wikipedia.org/wiki/Natural\\_language\\_processing](http://de.wikipedia.org/wiki/Natural_language_processing) (12.2007)
- 34) Wikipedia (2007): „Pattern Matching“: [http://de.wikipedia.org/wiki/Pattern\\_Matching](http://de.wikipedia.org/wiki/Pattern_Matching)  
(12.2007)
- 35) Wikipedia (2007): „Mustererkennung“: <http://de.wikipedia.org/wiki/Mustererkennung>  
(12.2007)
- 36) Wikipedia (2007): „Verborgenes Markow-Modell“:  
[http://de.wikipedia.org/wiki/Verborgenes\\_Markow-Modell](http://de.wikipedia.org/wiki/Verborgenes_Markow-Modell) (12.2007)
- 37) Wikipedia (2007): „Künstliche Intelligenz“: [http://de.wikipedia.org/wiki/K  
%C3%BCnstliche\\_Intelligenz](http://de.wikipedia.org/wiki/K%C3%BCnstliche_Intelligenz) (12.2007)
- 38) Wikipedia (2007): „Künstliche neuronale Netze“: [http://de.wikipedia.org/wiki/K  
%C3%BCnstliches\\_neuronales\\_Netz](http://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz) (12.2007)
- 39) Memex Technology Ltd. (): Memex Whitepapers: Abrufbar von  
<http://www.memex.co.uk/> (12.2007)
- 40) Institut für Geschichte, Salzburg (1999): „Big Brother is watching you“:  
<http://www.sbg.ac.at/ges/people/wagnleitner/sa/ms/msch.htm#Entwicklungen> (12.2007)