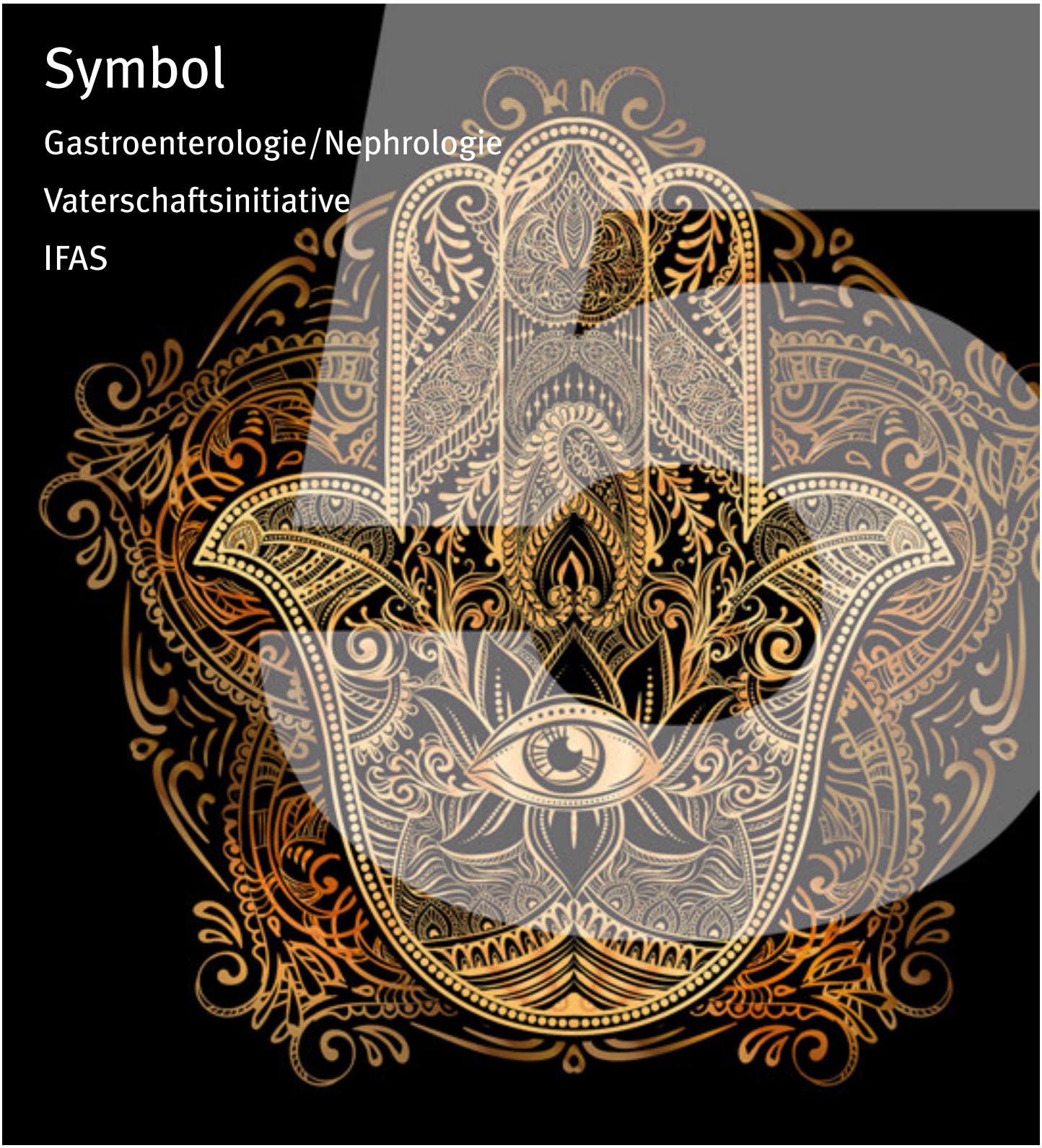
Nr. 5 Oktober 2016

VSAO ASMAC

VSAO JOURNAL



Symbol

Gastroenterologie/Nephrologie

Vaterschaftsinitiative

IFAS

Lebensversicherung für Daten

Sie kontrollieren, sie reparieren, sie sichern. Damit der Datenaustausch und die Speicherung von Daten problemlos verlaufen, braucht es Codes. Diese sorgen im Hintergrund dafür, dass z.B. die Präsentation nicht verloren geht, obgleich die CD zerkratzt ist. Die Kodierungstheorie sichert Daten gegen technische Probleme, aber auch gegen menschliches Versagen.

Dr. Anna-Lena Horlemann-Trautmann, Oberassistentin in Mathematik und Informatik an der EPF Lausanne

Heutzutage ist unser (berufliches wie auch privates) Leben ohne digitale Datenspeicherung und Kommunikation kaum noch vorstellbar – denken wir nur an Patientenakten oder Ultraschallbilder, die auf dem Computer gespeichert und bearbeitet werden, oder aber an E-Mails und Skype-Telefonate mit Freunden am anderen Ende der Welt.

Bei all diesen Technologien spielt die Kodierungstheorie eine Rolle. Diese hat die Aufgabe, die Datenspeicherung oder -übertragung sicher gegen technische Probleme bzw. physikalische Fehler zu machen. Solche Fehler können z.B. durch Stromausfälle in Rechenzentren, atmosphärische Störungen bei Funksignalen oder nicht mehr lesbare Speichermedien auftreten. In solchen Fällen möchten wir uns gegen den Datenverlust absichern – es wäre schliesslich tragisch, wenn komplette Patientenakten verloren gingen, nur weil eine Daten-CD verkratzt ist oder wenn eine E-Mail beim Empfänger nicht lesbar ist, weil ein Kabel zum Rechenzentrum des E-Mail-Providers beschädigt wurde.

Datenspeicherung

Die simpelste und intuitivste Art der Datensicherung ist die Sicherungskopie, also das Duplizieren der Daten. Wenn eine Datei teilweise nicht mehr lesbar ist, kann die Sicherungskopie herangezogen werden und davon wiederum eine neue Sicherungskopie gemacht werden. Die Informationen sind damit also nicht verloren gegangen. Bei einer solchen Sicherungskopie wird allerdings die doppelte Menge an Speicherplatz benötigt, bei einer doppelten Sicherungskopie dann schon der dreifache Speicherplatz usw. Bei grossen Datenmengen kann dies ein Problem werden. Die Kodierungstheorie beschäftigt sich unter anderem damit, wie man eine solche Absicherung gegen Datenverlust effizienter, also mit möglichst wenig benötigtem Speicherplatz, erreichen kann. Dabei werden die Daten vor dem Speichern «kodiert», indem man eine gewisse Redundanz hinzufügt. Mit dieser Zusatzinformation kann der Empfänger fehlerhafte Teile der Datei erkennen und korrigieren. Die simpelste Art von Redundanz ist das zuvor genannte Duplizieren der Nachricht, man kann allerdings durch fortschrittlichere Kodierungstechniken mit weniger Redundanz den gleichen Effekt erzielen.

Datenübertragung

Ein anderer Teil der Kodierungstheorie beschäftigt sich mit der Sicherheit von Datenübertragungen, also der digitalen Kommunikation. Ein Beispiel hierfür ist das Senden und Empfangen von E-Mails über WLAN. Bei der Datenübertragung in drahtlosen Netzwerken können schon kleine Störungen in der Luft die gesendeten Daten beschädigen und der Empfänger erhält somit nur einen Teil der gesen-

deten Nachricht korrekt. Um dieses Problem zu umgehen, wird der Inhalt der E-Mails, ähnlich wie bei der Datenspeicherung, kodiert, bevor die Nachricht gesendet wird. Der Empfänger kann dann die fehlerhaften Teile der Nachricht erkennen und korrigieren. Wie bei der Datenspeicherung ist es die Aufgabe der Kodierungstheoretiker, möglichst effiziente Kodierungen zu finden, also den Rohdaten möglichst wenig Redundanz für die gleiche Fehlerkorrektur hinzuzufügen. Dies ist wichtig, da jedes zu sendende Zeichen Energie verbraucht und gegebenenfalls eine Datenleitung für längere Zeit belegt.

Sehr viel schwieriger zu lösen als das Versenden von E-Mails ist ein kontinuierlicher Datenaustausch, z.B. bei einem Telefonat. Die Schwierigkeit ist hier, dass die Datenpakete live, also ohne Zeitverzögerung, beim Empfänger ankommen müssen. Im Gegensatz zum Dekodieren (also dem Korrigieren und Wiederherstellen) von E-Mails kommt hier erschwerend hinzu, dass die Fehlerkorrektur extrem schnell passieren muss. Falls dies z.B. bei Video-Telefonaten nicht funktioniert, kommt es zu den bekannten Aussetzern oder Verzögerungen in Bild oder Ton.

Beispiel Check Digits

Natürlich treten Fehler nicht nur durch technische Probleme und bei digitalen Daten auf, sie können ebenso durch menschliche Fehler und bei «analogen» Anwendungen entstehen.

Dass auch hierbei die Kodierungstheorie hilfreich ist, kann man anhand des klassischen Beispiels der sogenannten Check Digits sehen, welche z.B. bei Blutbeutel-Eurocodes oder Pharmazentralnummern (PZN) benutzt werden. Eine PZN besteht aus 8 Ziffern, wobei die Zahlen von 0 bis 9 und der Buchstabe X als Ziffer mit dem Wert 10 verwendet werden. Die ersten 7 Ziffern identifizieren das Medikament. Die letzte Ziffer, der sogenannte Check Digit, wird wie folgt aus den ersten 7 Ziffern berechnet: Man summiert die erste Ziffer plus zweimal die zweite Ziffer plus dreimal die dritte Ziffer usw. bis siebenmal die siebte Ziffer; dann teilt man diese Summe durch 11 und der Rest bei der Division ist der Check Digit.

Ist z.B. ein Medikament mit der Nummer 0116144 identifiziert, dann berechnet man zunächst die Summe: $1\times0+2\times1+3\times1+4\times6+5\times1+6\times4+7\times4 = 86$. Das Ergebnis durch 11 geteilt gibt einen Rest von 9, da $86 = 7\times11+9$. Damit ist der Check Digit die Ziffer 9 und die vollständige PZN des Medikaments ist 01161449.

Vertippt man sich beim Eingeben einer PZN, so dass eine der ersten 7 Ziffern falsch ist, dann ist die letzte Ziffer folglich nicht der Check Digit der ersten 7 Ziffern. Ein Computer kann auf diese Weise sofort feststellen, ob die eingetippte Nummer eine gültige PZN ist oder nicht. Falls im obigen Beispiel durch einen Fehler die zweite Ziffer von 1 auf 8 geändert wurde, erhalten wir den Check Digit 1, da $1\times0+2\times8+3\times1+4\times6+5\times1+6\times4+7\times4 = 100 = 9\times11+1$, entsprechend können wir direkt erkennen, dass 08161449 keine gültige PZN ist und somit ein Fehler vorliegt. Das Erkennen von falsch eingegebenen PZN ist nicht nur praktisch, sondern kann lebensrettend sein, da ohne die Check Digits Medikamente leicht vertauscht und somit falsch verabreicht werden könnten. Das Gleiche gilt für das Vertauschen von Blutbeuteln.

Check Digits gibt es aber nicht nur in der Medizin, sondern in vielen Bereichen unseres täglichen Lebens, z.B. bei Kontonummern, Banknoten, ISBN (Bücher), Fahrzeugnummern, Personalausweisen und vielen mehr.

Forschung

Neben der gerade dargestellten grundlegenden Technik von Check Digits gibt es in der Kodierungstheorie viele mathematisch fortgeschrittene Resultate zu Eigenschaften (wie Übertragungsrate und Fehlerkorrekturkapazität) von verschiedenen Codes und Algorithmen, wie diese möglichst effizient zu verwenden sind. Diese Resultate basieren zumeist auf Techniken und Resultaten aus der Algebra oder der Wahrscheinlichkeitstheorie.

Wegen der vielfältigen Anwendungen und der zugrunde liegenden mathematischen Theorie ist dieses Gebiet fächerübergreifend und sowohl Mathematiker wie auch Informatiker und Elektrotechniker befassen sich damit.

Kryptographie

Mit der Kodierungstheorie eng verwandt, und in der Öffentlichkeit mit mehr Interesse bedacht, ist die Kryptographie. Sie befasst sich auch mit Datensicherheit, allerdings geht es nicht um technische Probleme, sondern um das unerwünschte Abhören oder Abfangen von Daten bei der digitalen Kommunikation. Z.B. ist ein WLAN-Netzwerk, wie wir es zuhause benützen, recht ungeschützt: Mit ein wenig Know-how und einem normalen Laptop könnte jemand vor unserer Haustür alle gesendeten E-Mails abfangen und mitlesen, wenn sie nicht zuvor durch unsere E-Mail-Software kryptographisch verschlüsselt werden. Ähnlich ist es bei der Datenspeicherung, bei der wir Kryptographie benötigen, um Dateien mit einem Passwort (oder auch einem Fingerabdruck) zu verschlüsseln, so dass nur der Inhaber des Passwortes (bzw. des Fingerabdrucks) diese öffnen bzw. lesen kann. Obwohl die Aufgaben der Kryptographie und der Kodierungstheorie sich ähnlich anhören, sind die angewandten mathematischen Techniken in beiden Gebieten zumeist ganz andere. Für viele Anwendungen in der digitalen Kommunikation und Datenspeicherung sind allerdings beide Arten von Sicherheit äusserst wichtig, daher werden Kodierungstheorie und Kryptographie beim digitalen Datenaustausch vielfach in Kombination verwendet. ■