



Von Dipl.-Ing. Hans-Georg Wenke

Archive werden zu Datenquellen

Retrievalsystem erschließt den Fundus digital gespeicherte Texte- Software mit »natürlicher Intelligenz« und bahnbrechender Architektur

SOFTWARE

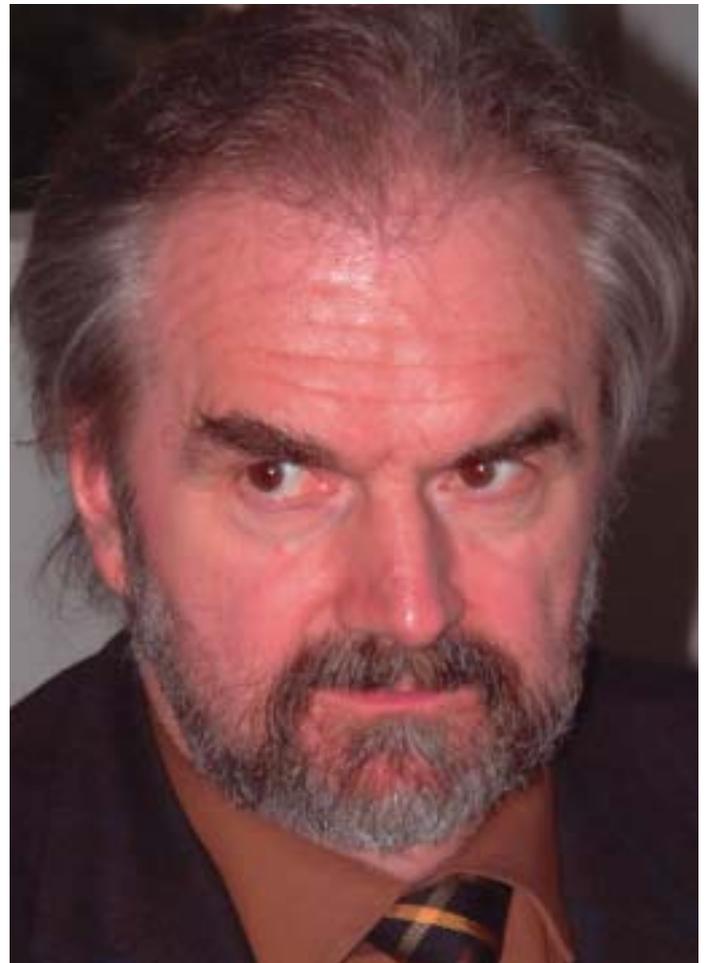


Das größte Dilemma der durch Computer verursachten Informationsflut ist die Verschwendung von Ressourcen. Informationen, die einmal in den Computer eingegeben wurden, ständen theoretisch aufgrund ihrer digitalen Natur ohne Verlust wieder zur Verfügung. Fakt ist jedoch, dass im Dschungel der Festplatten, Ordner und Datenformate selbst diejenigen keinen Durchblick mehr haben, die solche Daten selbst kreiert haben. Und von den Daten anderer Menschen (selbst im gleichen Unternehmen) weiß man so gut wie nichts, obwohl dies wichtige Informationen sein können. Weshalb die Datenverarbeitung insgesamt ihr eigenes Paradoxon geschaffen hat: Um den Reichtum elektronischer Archive zu nutzen, muss es Menschen geben, die dies alles im Kopf verarbeiten können.

Neueste Untersuchungen machen das Dilemma noch deutlicher und zeigen die Dramatik auf. Rund 70% aller Beschäftigten sind heute direkt oder indirekt von der Aufbereitung von Informationen und dem entsprechenden Zugang abhängig. Der Schaden, der durch nicht zugreifbare Informationen entsteht (fast gleichbedeutend mit Nichtwissen), kann durchaus mit einem Anteil von bis zu 10% der Arbeitsleistung quantifiziert werden. Darüber hinaus ist das Verknüpfen bestehenden Wissens eine qualifizierte Kreativität und unterstützt das »smart working«, also die intelligente Art, zu arbeiten.

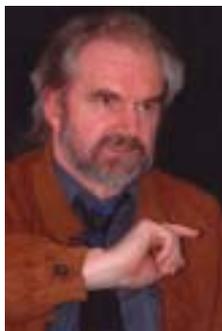
Wissensverlust ist existenziell

Der Verlust von Wissen kann existenziell gefährdend sein. »Knowledge Management« ist nicht von ungefähr eine der wichtigsten Aufgaben einer Unternehmensleitung. Wer das Wissen seiner Firma an die physikalische Präsenz von Menschenhirnen als Datenspeicher bindet, bekommt nicht selten massive Probleme: Fällt dieser Mensch aus, ist das Wissen verloren. Und das Rad muss erneut erfunden werden, Chancen werden verpasst, massive Fehler stellen sich ein. Dennoch wäre auch der umgekehrte Weg kein guter Rat, alles Wissen unkontrolliert in Computer zu stopfen. Manche Berufe oder Geschäftszweige sind sogar Informationsverarbeiter »pur«; in unserer Branche zählen dazu sämtliche Redaktionen und



Verlage. Sie leben von der Verbindung beider Methoden: »ahnende Menschen« brauchen »wissende Archive«. Dies ist wörtlich zu nehmen. Denn alle Kreativität der Menschen geht von Assoziationen aus. Kein Mensch »weiß« wörtlich genommen etwas, man »erinnert« sich allenfalls. Und Erinnerung ist immer Assoziation, nämlich das Verknüpfen von Begriffen, Bildern, Gegeben-

Der Schöpfer des Programmes, Manfred Hoffleisch, sagt ausdrücklich, dass es sich nicht um eine Erfindung in dem Sinne handelt, dass er etwas schuf, was es zuvor nicht gab. Er hat einen bestimmten universellen und damit »in der Natur vorkommenden« Algorithmus gefunden, der in der Lage ist – sehr vereinfacht ausgedrückt – Form und Inhalt zu trennen. Also Worte in Bedeutung und Form zu zerlegen.



Manfred Hoffleisch ist der »geistige Vater«, also Schöpfer der Grundidee, der darauf beruhenden Algorithmen, ihrer Umsetzung in Programme und Funktionalitäten. Die-ser Kern ist vielfältig anwendbar: er lässt unterschiedliche Anwendungen und Ausprägungen zu. Hoffleisch sprüht voller

Ideen, diese Möglichkeiten auszureizen und kann mit einer Reihe in Arbeit befindlicher Weiterentwicklungen aufwarten. Dem Sinn nach kann man es vergleichen mit einem Betriebssystem, das die unterschiedlichsten Applikationen zulässt oder sogar vereint. Und selbst bis in diesen Bereich, nämlich die Neudefinition von hardware-gebundenen Computerfunktionen gehen die von Hoffleisch entwickelten außergewöhnlichen Lösungen.

heiten, Gefühlen. Bei letzteren müssen Computer passen; die faktischen »Informations-Bits« können einem Computer sehr wohl in die Speicher gegeben werden.

Zu Wissen gesellt sich immer auch Logik

Bislang jedoch waren Programme, so sehr man sich auch an ihre Nutzung gewöhnt hat, im Bereich der Dokumentenverarbeitung alles andere als assoziativ. Sie waren im Gegenteil ausschließlich: Was nicht in einer Datei stand, war auch nicht als Wissen vorhanden.

Doch Wissen in dem Sinne, wie wir das Wort gebrauchen und Wissen, wie wir es als Werkzeug verwenden, ist etwas anderes als das Repetieren von Fakten. Zu Wissen gesellt sich nämlich immer die Logik. Und mit Logik haben Computerdateien wahrlich nichts am Hut. Man kann ihnen den größten Quatsch eingeben, sie interpretieren es als digitale Nullen und Einsen ebenso wie die Gedanken, Worte oder Texte von Genies.

Daraus ergibt sich ein Dilemma, wie es schlimmer nicht sein könnte: Wer etwas in digitalen Datenbanken sucht, muss wissen, wonach er sucht. Streng genommen geht es bei sogenannten Wissens-Datenbanken niemals wirklich um Wissen, sondern nur um das Finden von Speicherorten.

Normale Datenbanken sind nämlich »doof«: sie wiederholen, was man

ihnen gesagt hat. Und zwar differenzierungslos. Das lag und liegt unter anderem an den geradezu hölzernen Programmsprachen und Datenformaten, die sich, seit Computer erfunden wurden, eigentlich nicht geändert haben.

Hirne funktionieren unterschiedlich

Menschenhirne funktionieren völlig anders als Elektronengehirne. Sie haben den bislang unerreichten Vorteil, Wissen strukturiert speichern zu können. Indem sie Verknüpfungen herstellen und diese im Unterbewusstsein als Beziehungsgeflecht so anlegen, dass man sich der Erinnerung (also dem Wissen) von mehreren Seiten nähern kann.

Man könnte völlig unterschiedliche Begriffe nennen wie »hoch« oder »Schnee«, »Fels« und »Seilbahn«, »typisch Schweiz« oder »Wandertour«, und jedesmal könnten Menschen auf den Begriff »Berg« kommen. Es gibt »Eselsbrücken«, regelrechte Trainingsmethode, solche Begriffe an andere Begriffe zu knüpfen. Und der typische Effekt, dass plötzlich jede Menge Witze erzählt werden, wenn nur einer damit beginnt, beruht auf der gleichen Methode, eben der Assoziation. Das eine ergibt das andere. »Vom Hündchen auf das Stöckchen kommen«, lautet dazu ein Sprichwort.

Wer in herkömmlichen Datenbanken sucht, dem steht diese menschliche Methode nicht zur Verfügung. So

verrückt es klingt – wer im oben genannten Beispiel nicht weiß, dass er den Begriff »Berg« sucht, wird ihn nicht finden.

Überdimensionalen Sphären

Es sei denn, er benutzt HI-One, ein Text-Speicher-Management-System (Bilder sind wegen der Komplexität im Moment nicht einbezogen) einer ganze anderen, neuen, völlig einzigartigen Generation. Die dahinterstehende Programmierung ist so ihrer Zeit voraus, dass man nicht umhin kommt, Reizvokabel zu nennen, um sie zu beschreiben: Neuronale Netzwerke und Speicher, rekursive Algorithmen, Zusammenführung von Syntax, Semantik und Pragmatik mittels der Semiotik (Begriffserläuterung siehe Kasten auf der gegenüberliegenden Seite).

Das klingt nach überdimensionalen Sphären. Und wenn man ehrlich sein soll, ist es das auch ein wenig. Nämlich die Sphäre oberhalb der maschinellen »Wenn-dann-Mechanik« und die Brücke zur menschlichen Art, zu Denken.

Bislang konnten Programme so viel, wie die Programmierer ihnen mitgegeben hatten. Ihre Funktionalität war endlich, was nicht programmiert war, wussten und konnten sie nicht. Die selbstlernenden Programme sind zwar in der Informatik, Kybernetik und anderen Wissenschaften längst konkrete Vision und erklärtes Ziel. Alleine, sie zu schaffen ist bisher nur in Ansätzen gelungen.

Chromos hält den Atem an

HI-One ist eine auf Redaktionen und Verlage abgestimmte Programm-Variante (die aus verschiedenen Modulen besteht) der von HumanInterfaces (HI) entwickelten »neuronalen Programmstruktur«. In der Schweiz werden die Anwendungen von der Chromos AG, Glattbrugg (ZH), vertrieben. Im bekannten blauen Chromos-Haus ist HI seit kurzem als Mieter zu Gast, schweizer Fachleute und Programmierer implementieren die Software und garantieren den Support in Verbindung mit Chromos.



»Wir halten selbst den Atem an, jetzt, nachdem wir wissen, was mit dieser Art der Programme möglich ist«, sagt Klaus Kraemer, Leiter des Chromos-Unternehmensbereiches PrePress. Er hofft zusammen mit Daniel Mäder, der Interessierten die HI-One-Lösung im Detail erläutert und bei der Implementierung berät, auf »aufgeschlossene Unternehmen, die die Schätze ihren Mitarbeitern zugänglich machen wollen, die in den jeweiligen digitalen Archiven oft völlig ungenutzt ruhen«.

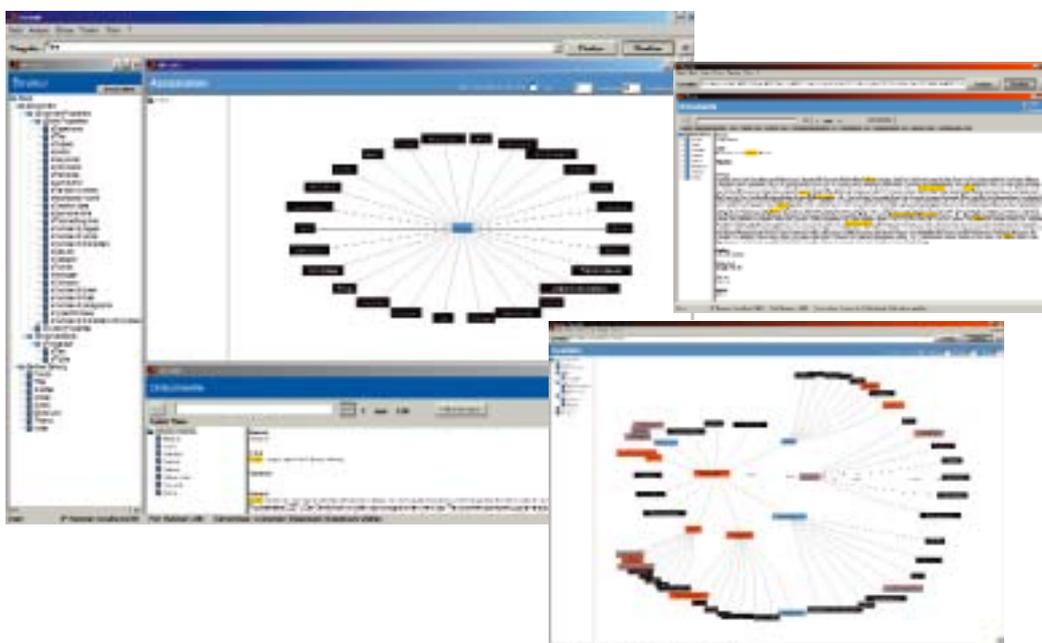
Die Kosten des Programms sind, gemessen an seiner Effizienz, absolut im grünen Bereich. Der Aufwand, den es erfordert, rentiert sich extrem schnell gemessen an den Zeitaufwendungen, die bei Nichtnutzung gespeicherten Wissens erforderlich ist. Hinzu kommt der verstärkende Effekt, den Daniel Mäder sehr einfach formulieren kann: »Je mehr Texte und Dokumente, je größer das Archiv, desto höher der Nutzen.«

Die Kosten des Programms sind, gemessen an seiner Effizienz, absolut im grünen Bereich. Der Aufwand, den es erfordert, rentiert sich extrem schnell gemessen an den Zeitaufwendungen, die bei Nichtnutzung gespeicherten Wissens erforderlich ist. Hinzu kommt der verstärkende Effekt, den Daniel Mäder sehr einfach formulieren kann: »Je mehr Texte und Dokumente, je größer das Archiv, desto höher der Nutzen.«

Druckmarkt macht sein Wissen zugänglich

Für den 6. März hat die Chromos AG ab 10.00 Uhr Interessierte eingeladen, das Programm und seine Möglichkeiten selbst in Augenschein zu nehmen und ausführlich zu testen.

Im Rahmen der Programmdemonstration werden Interessierte tief in den Archiven des Druckmarktes wühlen können. Wir stellen der Chromos AG und HumanInterface speziell für diesen Zweck unser gesamtes nutzbares Dokumentenarchiv zur Verfügung. Sie werden also die Recherchen mit Branchen-vokabular durchführen können und brauchen nicht auf abstrakte Beispiele auszuweichen.



Je umfangreicher die Datenbank, desto imposanter werden die Fähigkeiten des Programms deutlich. Die Lösung sucht und findet nicht nur gesuchte Schlagworte, sondern stellt Beziehungen und Zusammenhänge her. Gibt man einen expliziten Begriff ein, schlägt einem das Programm (in einstellbarer Anzahl) assoziative Begriffe vor. Zu »Hund« findet das Programm mindestens auch »Leine«, »Fressnapf« oder »beißen« (sofern diese Worte in beliebigen Dokumenten vorkamen). Es findet Zusammenhänge, die man nicht wusste, es zeigt Verbindungen auf, die einem helfen, kreativ zu sein. Es führt zu Quellen, die man ohne solch ein Programm nicht gefunden hätte.

GLOSSAR

Linguistik gesamter Bereich der Sprachwissenschaften.

Pragmatik Sinn für das Machbare, für Tatsachen, Orientierung auf das Sinnvolle.

pragmatisch Sinn für das Nützliche, Sinnvolle oder Machbare; sich auf das Naheliegendste beschränkend, realistisch.

Semantik Teilgebiet der Linguistik, das sich mit der grundsätzlichen Analyse und Beschreibung der Bedeutung sprachlicher Ausdrücke befasst.

semantisch 1. Bedeutung eines sprachlichen Ausdrucks betreffend; 2. Die > Semantik betreffend, auf ihr beruhend, zu ihr gehörend.

Semiotik allgemeine Theorie der sprachlichen und nichtsprachlichen Zeichensysteme.

semiotisch 1. das Zeichen betreffend; 2. die > Semiotik betreffend.

Syntaktik Teilgebiet der > Semiotik, das sich mit den formalen Beziehungen zwischen sprachlichen Zeichen befasst.

Syntax 1. Regelsystem einer Sprache (> Linguistik); 2. regelgerechter Satzbau; 3. formale Regeln einer Programmiersprache in der EDV.

Wenn nun hier die Rede davon ist, dass ein solches Programm ab sofort zu sehr überschaubaren finanziellen Konditionen praktisch jedermann zur Verfügung steht, ist es nicht verwunderlich, dass dies auf Skepsis

stößt. Was die wirklich Interessierten nicht davon abhalten sollte, zu erforschen, was nun real und wahr ist.

Findung, nicht Erfindung

Das Geheimnis dieses Programmes liegt in einer »Findung«. Der Schöpfer Manfred Hoffleisch sagt ausdrücklich, dass es sich nicht um eine Erfindung in dem Sinne handelt, dass er etwas schuf, was es zuvor nicht gab. Er hat einen bestimmten universellen und damit »in der Natur vorkommenden« Algorithmus gefunden, der in der Lage ist – sehr vereinfacht ausgedrückt – Form und Inhalt zu trennen. Also Worte in Bedeutung und Form zu zerlegen. Und die Bedeutungen so zu programmieren, dass sie in der Lage sind, zu anderen Bedeutungen »Verbindungen« aufzunehmen und »Klassen« zu bilden. Verbindungen, die sogar erst dann aktiviert werden, wenn sie von bestimmten Impulsen »gereizt« werden. Im Prinzip ist es nicht mehr oder weniger als die Selbstorganisation der belebten Natur.

Es ist logisch, dass es sich in dem nunmehr in Programmen nutzbar gemachten Algorithmus keineswegs um Esoterik oder gar die »Weltformel« handelt, nach der ganze Scharen seriöse Wissenschaftler zu allen Zeiten hinterher waren und sind. Nein, es geht um eine Art der Informationsverarbeitung und Computerprogramme, die strukturell mit der Denk-, Handlungs- und Entschei-

dungsweise der Menschen besser zusammenarbeiten.

Jeder Mensch, so Manfred Hoffleisch (www.hi.ag) speichert Wissen und betrachtet bzw. beurteilt die Welt in einem Modell der Welt. Dies könnte man auch Vorurteil nennen; es ist eben jene Betrachtungsweise, die individuell ist und einschließt, das sich manche engstirniger als andere anstellen.

Die Programme, die Hoffleisch und seine Partner geschrieben haben, entkleiden (Text-)Informationen von der Individualität der persönlichen Betroffenheit und setzen die Texte der ihnen innewohnenden Bedeutung ihrer Inhalte nach in Beziehung. Dabei können Menschen »editierend«, gewissermaßen lernend-lehrend eingreifen und dem Programm durch Strukturen und konkrete Zusammenhänge »auf die Sprünge helfen«.

Der Effekt ist absolut verblüffend, die meisten, die es zum ersten Mal sehen, mögen es nicht glauben und denken, dies wäre ein gut vorbereiteter »Fake«, wie es in der Filmsprache heißt, eine Vortäuschung. Ist es aber nicht.

Man gebe »jede Menge Texte« (je mehr, desto besser) mit Hilfe des Programms in eine »Datenbank« ein. Das Programm sorgt selbst für die Aufbereitung in der Form, dass etwas zu Stande kommt, was Rechercheure absolut zu schätzen wissen. Man kann nämlich »unscharfe« Suchen starten, mit assoziativen Ahnungen beginnen, mit

den Begriffen, Namen, Substantiven, die man selbst mit dem Suchbegriff in Verbindung bringen würde (Assoziationskette). Und umgekehrt: gibt man einen expliziten Begriff ein, schlägt einem das Programm (in einstellbarer Anzahl) assoziative Begriffe vor. Fast schon persiflierend platt konkretisiert: Zu »Hund« findet das Programm mindestens auch »Leine«, »Fressnapf« oder »beißen« (sofern diese Worte in beliebigen Dokumenten, und das ist das absolut Revolutionäre, vorkamen). Es findet Zusammenhänge, die man nicht wusste, es zeigt Verbindungen auf, die einem helfen, kreativ zu sein. Es führt zu Quellen, die man ohne solch ein Programm nicht gefunden hätte.

Datamining

Das ist, was mit dem Begriff »data mining« im eigentlichen Sinne gemeint ist. Das »Schürfen« nach Informationen und Daten, von denen man allenfalls hoffte, sie zu finden, aber nicht wusste, wo sie waren. Das Auswerten von Archiven (jede Sammlung eingegebener Informationen, egal welcher Art, und undabhängig, ob sie zusammengehören oder nicht, ergibt ein Archiv) wird so zur Wissens- und Informationsvernetzung einer neuen Qualität. Nämlich der, dass jeder die gleiche Chance hat, auf Informationen zu kommen, die stimmig sind und die ohne Vorwissen darüber zu haben dennoch einen extrem hohen Grad logischer Verbindung aufweisen.

HI-One ist ein dynamisches Informations-Management-System der Human Interfaces AG. Die von dem Unternehmen entwickelte Software ermöglicht es erstmals, ähnlich dem menschlichen Gehirn begriffliche Zusammenhänge zwischen gespeicherten Daten herzustellen. Damit wird ein Computer mit Fähigkeiten ausgestattet, die mit der menschlichen Intelligenz fast vergleichbar ist.



- Die Software ermöglicht die semantische Erfassung, Verarbeitung und Auswertung von strukturierten und unstrukturierten Texten (Dokumenten; Quark-Files, PDF-Files, XML-Files; Zahlen, z. B. auch Excel-Sheets; das Programm wandelt automatisch in XML-Strukturen um).
- Das Einlesen und Einbeziehen von Internet-Domains (mit einstellbarer Link-Hierarchie und Verschachtelungstiefe) ist möglich.
- Das Programm verbindet automatisch Begriffe und deren Bedeutung nach einer dem menschlichen Gehirn ähnlichen Struktur und vom Menschen nachvollziehbarer Logik.
- Die Suche führt auf in XML-Form gespeicherte Texte, die auf ASCII-Basis wieder in Dokumente übernommen werden können.
- Die Lösung beinhaltet vorerst zwei Programmmodule: Stellaris als Datenspeicher und ULF als Universal Language Flow für die Zusammenhänge. Der spätere Ausbau mit Einbeziehung anderer Strukturen und Elemente setzt voraus, die Text-Basis zu beherrschen (schrittweiser Einstieg unbedingt empfohlen).
- Die Suche nach Texten, Worten, Begriffen vollzieht sich im Bereich weniger Sekunden, auch bei einer sehr großen Datenmenge.
- Die assoziativen Wörter bzw. Fundstellen sind gemessen an den Date Ursprüngen dokumenten- und file-übergreifend und völlig format-unabhängig; das heißt es spielt keine Rolle mehr, welches Dateiformat die Texte vorher hatten.
- Das Archiv ist anwendbar für alle Arten von Daten, Informationen, Dokumenten, Texte, Sprachen, Zahlen. Auch gemischte Inhalte werden ohne Qualitäts- und Zeitverlust assoziativ verarbeitet.
- Einsatzbereiche sind Verlage, Redaktionen, Marktforschung, Wissenschaft; aber auch Produktionsumgebungen mit komplexen File- und Jobstrukturen (PrePress, Druckerei, Agentur, IT-Abteilungen usw.).
- Auch für den professionellen Einsatz sind lediglich Normal-PCs erforderlich; schnelle und leistungsfähige, aber keineswegs spezielle PCs oder Server.
- Die Software ermöglicht jegliche Form von »intelligentem Wissensmanagement« durch freie Sucheingaben, offene Nutzung und minimalem Administrationsaufwand.
- Die Verarbeitung völlig unstrukturierter Daten und Dokumente ist möglich, aber auch die freie Definition von Strukturen (auch nachträglich jederzeit); durch Strukturen erhöht sich die Treffer- und Assoziationsquote um messbare Dimensionen.

alle Datenbank-, Archiv-, Retrieval- ein Programm von der Sorte, die (Auffindung), Knowledge- und Redaktions- oder Dokumentenmanagementsysteme bisher waren. Eben man als epochal bezeichnen darf.

➤ www.chromos.ch

➤ www.hi.ag

Damit verschafft HI-One den Zeitungs- und Verlagsarchiven den Eintritt in eine Cyberwelt, die anders als befürchtet extrem menschlich-normale Strukturen hat. Aber auch das Wissensmanagement in größeren Unternehmungen, erst recht Konzernen, kann in einer Art und Weise so effizient gestaltet werden, wie es ohne Hilfe der IT-Strukturen (also nur durch personengebundenen Wissenstransfer) niemals hätte sein können. Wenn nicht »das Wissen der Welt«, so kann aber dennoch jeder Mitarbeiter, der über die Suchsoftware verfügt, das »Wissen des Unternehmens« für seine jeweiligen expliziten Belange höchst kreativ nutzen. Die Querverbindungen, das ergeben Tests immer wieder, decken Fakten und Zusammenhänge auf, die vorher nicht bewusst waren.

Ungeahnte Möglichkeiten

Insofern fördert die Software Wissen zu Tage, von dem man vorher gar nicht wusste, dass es von Bedeutung ist. Sie bringt einen, wörtlich genommen, auf völlig neue, aber immer realitäts-, sprich text- und zahlenbezogene Ideen. Was das für Wis-

senschaft und Forschung, Lehre und Konstruktion sowie artverwandte Disziplinen bedeutet, mag man ahnen: Es lassen sich Aufgaben lösen, die bislang ungelöst bleiben mussten.

Das Programm und die dahinterstehende Philosophie kennenzulernen, ist – wollte man nichts weiteres bezwecken – ein spannendes geistiges Abenteuer. Kommt man aber aus dem Metier der Verlage und Redaktionen (Zeitungen, Zeitschriften, Bücher, Sammelwerke; die Disziplinen, Fakultäten und Themen spielen keine Rolle) oder aus dem Bereich Marktforschung und Marketing, dann ist HI-One mehr als Kennenlernen wert.

Epochale Entwicklung

Es könnte sich, konsequent eingesetzt, zu einem Tool entwickeln, das die Arbeit und die Art, Wissen zu finden, zu nutzen, zu vermarkten und zu fördern revolutioniert. Diese Extremvokabel muss erlaubt sein, da das Programm bereits in seiner jetzigen ersten Form (der Ausbau ist realisiert, bedingt aber das Etablieren der Einstiegsstufe) anders ist, als

**Lesen Sie nicht alles Mögliche.
Lesen Sie das Wesentliche!**

Druckmarkt Schweiz ■ Themen auf den Punkt gebracht

