

# Typische Künstliche Intelligenz-Lösungen

## "mit künstlichen Neuronalen Netzen"

Künstliche Neuronale Netze (KNN) sind ein Verfahren der Künstlichen Intelligenz, das versucht, Lernprozesse des menschlichen Gehirns nachzubilden. KNNs eignen sich dabei besonders für Mustererkennung und Klassifizierung. HMI hat hieraus ein Softwaresystem abgeleitet, das effizient Dokumente klassifiziert, egal ob es sich hierbei um Office Dokumente, Faxe, PDFs, E-Mails oder sonstige digitale Texte handelt. Komplexe statistische Verfahren und schnelle mathematische Bibliotheken unterstützen dabei den Prozess, sowohl auf der Ein- als auch auf der Ausgabenseite.

*Kommentar: Traditionelle "KNN" berechnen inhaltliche Bedeutung in der Regel mittels Vektor Parametern. D.h. bei der Indizierung wird die Distanz und Vernetzung einzelner Worte gemessen und es werden Winkel und Strukturen gebildet. Mit dieser Methode erreicht man ca. eine 40%ige Trefferquote, verglichen mit der menschlichen Referenz. Ein Problem dieser Methode ist, dass inkrementelle (fortlaufende) Updates nur bedingt bis gar nicht möglich sind. Verändert sich der Daten Korpus, muss immer wieder neu von Anfang indexiert werden.*

**ai-one™ (BII) biologisch inspirierte Intelligenz, verfolgt einen neuen Ansatz. Der holosemantische Data Space ist eine neue Generation der neuronalen Technologie. Er arbeitet vollständig ohne Statistik oder vordefinierte Gewichte und Regeln. Der holosemantische Data Space ist ein zellulärer Raum, welcher sich vollständig selbst organisiert, eine innere Motivationsunruhe hat und zugleich selbstlernend ist. Für Sprache bedeutet dies, dass die inhärente Semantik und alle semiotischen Muster in einer Sprache, welche z.B. in Text-Form (PDF, Mail; Brief etc.) vorliegen, erkannt werden. Inhärente Semantik heisst, die innewohnende Bedeutung des gesamten Textes und der Worte. Die semantische Trefferquote beim ai-one™ Verfahren liegt doppelt so hoch wie beim konventionellen Verfahren und es kann jederzeit ohne neue Berechnung der Semantik inkrementell mit Daten erweitert werden.**

Genetische Algorithmen sind heuristische Verfahren aus dem Gebiet der Künstlichen Intelligenz. Prozesse aus der Evolution wie Selektion, Mutation und ‚survival of the fittest‘ werden hierbei auf ein gegebenes Problem angewandt. Es werden Individuen in Gruppen zusammen gefasst, deren als ‚Gene‘ bezeichnete Parameter sich über die Generationen hinweg verändern. Hierdurch findet eine Optimierung der Parametersätze statt. Genetische Algorithmen eignen sich besonders gut für komplexe Probleme mit einer großen Anzahl von Parametern, die sich nicht oder nur schwer durch konventionelle Ansätze lösen lassen.

*Kommentar: Das sind gute und aktuelle Verfahren. Das Problem ist, dass diese Verfahren zu spät angewandt werden. Wenn der Text bereits in ein "KNN" importiert und indexiert wurde, ist die original inhärent vorhandene Semantik bereits verändert worden. Es bleibt nur noch eine, nach linguistischen Regeln vordefinierte und vorklassifizierte Semantik übrig.*

**Der ai-one™ Algorithmus, genannt OPTI-Algorithmus unterscheidet sich ebenfalls von den traditionellen genetischen Algorithmen. Er entdeckt zwingend jegliche Informationspattern, ohne dass er wissen muss, worauf er achten soll. Eigenschaften wie Evolution, Mutation und Überleben des Stärkeren, sind natürliche Eigenschaften des OPTI-Algorithmus. Darüber hinaus kann der ai-one™ Algorithmus aber auch zukünftige Ereignisse mit einer Wahrscheinlichkeitsbewertung vorhersagen. Er kann komplexe und verschachtelte Muster und Strukturen aufdecken (erkennen) und ihre Bedeutung im Kontext über numerische Werte bewerten.**

Dank einer Kombination aus innovativen linguistischen Programmierungsansätzen mit komplexen Statistikmodulen sind unsere Produkte in der Lage, auf eigenständige Weise und ohne manuelle Regeldefinitionen durch den Menschen, im Hintergrund zu lernen. So sind sie in der Lage, sich gezielt und zeitnah den jeweiligen Bedürfnissen und Anforderungen des Arbeitsumfeldes anzupassen. Das bedeutet, dass durch den Einsatz genetischer Algorithmen die Parameter des Systems kontinuierlich verbessert werden. Dadurch tragen sie zu einer genaueren Auswahl der Künstlichen Neuronalen Netze bei, die für Suchvorgänge innerhalb von Wissens- und Logikdatenbanken verwendet werden.

*Kommentar: Statistik und andere "boolesche" Verfahren, sowie berechnete Modelle sind ungeeignet, um eine dynamische selbstregierende Lösung zu erstellen. Linguistische Programmieransätze sind in der Regel linguistische Modelle und verwenden Ontologie, Thesauri und Taxonomien. Die Qualität der semantischen Grundfunktion definiert sich nur über die Qualität dieser vom Programmierer oder Linguisten vordefinierten Filter. Generische Algorithmen können natürlich jetzt eingesetzt werden, die Frage ist einfach, wie viel Qualität aufgrund der Vorbehandlung bereits unwiderruflich verloren ist.*

**ai-one™ geht da andere Wege mit dem holosemantischen Data Space. Dieser Data Space ermöglicht, dass die intrinsische Semantik und Semiotik bereits ohne Zutun von Programmierern oder Expertenwissen im Data Space abgebildet wird. Der Experte oder Programmierer kann jetzt die Resultate der inhärenten Semantik abrufen, auswerten und bearbeiten. Zusätzlich offeriert ai-one ein API-Interface „BrainBoard“ über welches der User einfach und schnell einen Rapid-Prototyp für unterschiedliche Problemstellungen und Lösungsansätze erstellen kann.**

## Vergleichstest

Damit die oben genannten Aussagen etwas besser verstanden und quantifiziert werden können, zeigt der nachfolgende Vergleichstest auf, wo sich die Unterschiede genau zwischen ai-one und der traditionellen KNN Technologie auswirken.

Dazu vergleichen wir den ai-one - Algorithmus mit dem cos/tf-idf - Algorithmus. Beide Verfahren ermitteln in unstrukturierten Texten (Prosa) die semantisch, assoziativen Beziehungen zwischen Worten. Sie erkennen (ai-one) bzw. berechnen (cos/tf-idf) die inhaltliche Nähe und Bedeutung der Worte zueinander.

## Resultat-Tabelle

Als Test-Menge wurden 200 Dokumente aus dem Thema Alkoholsucht gewählt. Mit entscheidend ist, dass eine semantische Analyse bereits aus einer kleinen Menge an Dokumenten gemacht werden kann, da oft nicht mehr an „Domain-Expertise“ Fachwissen zu Verfügung steht.

### Zusammenfassung des Tests:

Thema	ai-one	cos/tf-idf
1. Import / Indexierung	n.a.	~ 20hrs
1. Import / Semtische Analyse	~ 15min	n.a.
Inkrementeller Import (update)	JA	NEIN
2. Import plus eine (1) Seite	~ 30sec	~ 20hrs
Anzahl semantische Wort-Verbindungen	hoch/ Ø 53	tief/ Ø 34
Erwartete Treffer aus 100	Ø 73	Ø 43

Test-Frage	c = 0.8						c = 0.9					
	ai-one™			cos/tf-idf			ai-one™			cos/tf-idf		
	R	T	R/T	R	T	R/T	R	T	R/T	R	T	R/T
alcohol dependency	42	64	0.66	4	6	0.66	40	57	0.7	1	2	0.5
alcohol problem	44	64	0.69	7	16	0.44	10	16	0.63	1	4	0.25
prevention	13	20	0.65	0	0	0	12	17	0.7	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
therapy	79	116	0.68	9	13	0.69	71	104	0.68	0	0	0
treatment facility	6	8	0.75	10	48	0.21	4	6	0.67	10	46	0.22
withdrawal symptoms	29	32	0.9	15	34	0.44	28	31	0.9	6	9	0.67
<i>average</i>		60.9	0.73		38.3	0.47		46.6	0.73		30.8	0.41

Der Test zeigt einerseits auf, dass sich ai-one viel besser semantisch vernetzt und mehr Treffer liefert, als auch dass ai-one ohne Probleme fortlaufend updaten kann, wogegen herkömmliche Verfahren ein grosses Problem mit stetigem updaten haben und auch qualitativ/quantitativ weniger Treffen liefern.

## Was heisst das Testresultat für die Praxis

Einerseits zeigt es positiv auf, dass es möglich ist, semantische Beziehungen in schon relativ kleinen Dokument-Mengen zu erkennen. Es zeigt aber auch, dass es qualitativ und quantitativ grosse Unterschiede zwischen den verfahren gibt die sich wie folgt in der Anwendung / Praxis auswirken:

**In der Praxis sind sehr oft folgende Eigenschaften von semantischen Systemen gefordert, damit diese effizient, verlässlich und kostengünstig eingesetzt werden können:**

### Zusammenfassung des Tests:

Thema	ai-one	cos/tf-idf
Sprachunabhängig	JA	NEIN
Effizient	JA	Bedingt
Schnell	JA	NEIN
Hohe Qualität	JA	Bedingt
Spontan	JA	NEIN
Selbstlernend	JA	JA
Skalierbar	JA	Bedingt
Kostengünstig in der Anschaffung	Lizenzabkommen	Open Source
Kostengünstig in der Betreuung / Unterhalt	JA	NEIN

**Damit ist klar, der Einsatz vom cos/tf-idf kann in diversen Spezialapplikationen optimal sein. Es dürfen aber keine zeitkritischen und sich stetig verändernden Daten betreut werden müssen.**

ai-one hingegen kann sehr universell eingesetzt werden. Natürlich können die Datenmengen auch bei ai-one schnell sehr gross werden. Da ai-one aber über mehrere Instanzen und Cloud-Computing verteilt werden kann, können grosse Datenmengen trotzdem verarbeitet werden. (Skalieren)

## ai-one™ „Import“ bedeutet: semantische Analyse!

Wenn ai-one Dokumente „importiert“ ist das nicht dasselbe wie bei anderen Programmen oder Datenbanken! Diese anderen Programme indexieren einfach Worte und das geht relativ schnell...

**ai-one hingegen „liest“ alle Dokumente durch, findet alle semantisch, assoziativen Verbindungen zwischen den Worten. Brainup erkennt also die Inhaltliche Bedeutung zwischen den Worten und Sätzen und merkt sich dieses Wissen. Daher dauert ein „importieren“ bei ai-one länger als bei einer konventionellen Datenbank oder Suchmaschine, aber schneller wie mit jedem bekannten andere semantischen verfahren!**