

Patentrecherchen und Künstliche Intelligenz

Alexander Winninger

Serviva Austria GmbH

Die technologische Revolution, die wir seit einem merklichen Qualitätssprung der Künstlichen Intelligenz (KI) vor wenigen Jahren verzeichnen, betrifft unseren Umgang mit Informationsgewinnung und –einschätzung fundamental.

Es ist zur klassischen Suche nach technischen Inhalten, die auf deren Originaltext und dessen Indizierungen beruht, eine praktikable Variante hinzugetreten, die diese Dokumente als Gesamtheit indiziert und einschätzt, und somit auf besonders niederschwellige Art auffindbar macht – durch detaillierten Ähnlichkeitsvergleich von Texten, Abbildungen und auch ganzen Dokumenten.

Diese neue Methode bereichert das Gebiet des Suchens, Findens und Einschätzens von technischen Informationen ungemein und ist als solche praktisch nicht mehr wegzudenken, da sie die klassische Suche nicht simuliert, sondern eben als genuin neue Methode erweitert.

Die Hoffnung und Erwartung liegt nahe, dass der rasche, zielgerichtete und selbstlernende Algorithmus, den KI darstellt, die letztlich unüberblickbare Datenbasis technischer Dokumente effizient und möglichst umfassend zugänglich und damit optimal, ja gleichsam restlos, nutzbar macht.

Es liegt allerdings auf der Hand, dass die Methode der KI in sich, also systematisch, mit Fehlern behaftet ist, die mit dieser Methode selbst nicht erkannt oder behoben werden können. Ein Optimum von ihr zu erwarten, ist also eine Illusion, der man sich wahrscheinlich umso weniger hingeben möchte, je wichtiger die Sache ist, um die es geht.

Das Ergebnis einer Suche mit KI ist immer das Ergebnis einer Statistik auf der Basis von fremden Inhalten und Indizierungen, deren Bedeutungen von der KI nicht verstanden werden können. Der systematische Fehler dieser Methode sind die menschlichen Eingaben und Angaben, die dieser Statistik zugrunde liegen, die von der Methode selbst aber nicht überprüft werden können, da ihr das inhaltliche Verständnis fehlt, das sie auch nicht erwerben kann.

Positive Ergebnisse sind inhaltlich im Nachgang in Bezug auf ihre Relevanz leicht überprüfbar. Die Vollständigkeit einer Suche hingegen zeigt sich nicht in diesem positiven Resultat, die bleibt im Dunkeln.

Exakt aus diesem einen Grund will man wahrscheinlich, ganz besonders, wenn es wichtig ist, die klassischen Ansätze, Dokumente aufzufinden, also die menschliche Intelligenz, Erfahrung und Fertigkeit, genauso nutzen wie KI.

Der Mensch sucht unabhängig und ergänzt das maschinelle Resultat und validiert es, man verwendet die KI als Zusatz, nicht als Basis, weil man sich sonst ihre Brille aufsetzt und die Fehler systematisch zementiert.

Es sind verschiedene Methoden und man will sie insgesamt optimal nutzen – wenn es wichtig ist. Wenn eine Tendenz als Ergebnis reicht, nimmt man wahrscheinlich nur eine der Methoden, will man mögliche Vollständigkeit und Aussagekraft, bedarf es beider.

Diese allgemeinen Überlegungen werden durch die Beobachtung gestützt, dass im Schnitt bei Nichtigkeitsrecherchen die menschliche Methode die KI in allen Relevanzgruppen schlägt – also jeweils mehr relevante Dokumente findet, die die KI nicht gefunden hat, als umgekehrt – und sogar bei kurzen Recherchen zu Erfindungsmeldungen noch im Bereich der höchstrelevanten Dokumente (exemplarische Auszüge dieser Beobachtungen siehe Anhang A).

Mein Fazit: Das Feld der Suche nach technischen Informationen ist durch KI enorm bereichert worden – auf den Menschen wird man dennoch unabhängig nicht verzichten wollen oder können. Eben die voneinander unabhängige Nutzung der beiden Methoden erscheint in Summe besonders erfolgversprechend, da jeweils Ergebnisse erzielt werden, die mit der anderen Methode nicht oder nur schwierig zu erhalten sind.

Anhang A

Exemplarische Auszüge aus den Beobachtungen Mensch vs. KI: nur der Mensch bedeutet, dass die KI diese Dokumente nicht gefunden hat, nur die KI bedeutet, dass der Mensch diese Dokumente nicht gefunden hat; beide bedeutet, dass sowohl der Mensch als auch die KI diese Dokumente gefunden hat; alle angegebenen Dokumente wurden als zitierwürdig eingeschätzt; mittlere Relevanz bedeutet gute As, hohe Relevanz bedeutet Y, höchste Relevanz bedeutet X und ausgezeichnete Y; für die menschliche Suche wurde die Datenbank PatBase, als KI wurde IPRally verwendet (die Auswertung erfolgte ebenfalls in PatBase).

einzelne Nichtigkeitsrecherchen:

Gebiet	Verfahrens- technik				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	10	5	6	11	21
nur der Mensch	70,0 %	60,0 %	33,3 %	45,5 %	57,1 %
	7	3	2	5	12
beide	1	2	3	5	6
nur die KI	2		1	1	3
	20,0 %	0,0 %	16,7 %	9,1 %	14,3 %

Gebiet	Maschinen- bau				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	6	7	6	13	19
nur der Mensch	16,7 %	42,9 %	33,3 %	38,5 %	31,6 %
	1	3	2	5	6
beide			1	1	1
nur die KI	5	4	3	7	12
	83,3 %	57,1 %	50,0 %	53,8 %	63,2 %

Gebiet	Elektro- technik				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	8	9	5	14	22
nur der Mensch	50,0 %	33,3 %	60,0 %	42,9 %	45,5 %
	4	3	3	6	10
beide	3	1		1	4
nur die KI	1	5	2	7	8
	12,5 %	55,6 %	40,0 %	50,0 %	36,4 %

Mittelwerte (über 6 Nichtigkeitsrecherchen):

Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
nur der Mensch	54,9 %	56,6 %	58,9 %	56,5 %	56,0 %
nur die KI	36,0 %	34,9 %	28,3 %	32,6 %	34,4 %

einzelne Recherchen zu Erfindungsmeldungen:

Gebiet	Maschinenbau				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	1	2	2	4	5
nur der Mensch	100,0 %	50,0 %	50,0 %	50,0 %	60,0 %
beide	1	1	1	2	3
nur die KI	0,0 %	50,0 %	50,0 %	50,0 %	40,0 %

Maschinenbau				
mittel	hoch	höchst	hoch/höchst	gesamt
6	5	4	9	15
33,3 %	40,0 %	25,0 %	33,3 %	33,3 %
2	2	1	3	5
	1	1	2	2
4	2	2	4	8
66,7 %	40,0 %	50,0 %	44,4 %	53,3 %

Gebiet	Elektrotechnik				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	4	8	2	10	14
nur der Mensch	0,0 %	0,0 %	100,0 %	20,0 %	14,3 %
beide	3	3	2	3	6
nur die KI	1	5	0,0 %	5	6
	25,0 %	62,5 %	0,0 %	50,0 %	42,9 %

Elektrotechnik				
mittel	hoch	höchst	hoch/höchst	gesamt
6	6	6	12	18
33,3 %	50,0 %	66,7 %	58,3 %	50,0 %
2	3	4	7	9
2	2		2	4
2	1	2	3	5
33,3 %	16,7 %	33,3 %	25,0 %	27,8 %

Gebiet	Elektrotechnik/Verfahrenstechnik				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	4	6	3	9	13
nur der Mensch	50,0 %	33,3 %	33,3 %	33,3 %	38,5 %
beide	2	2	1	3	5
nur die KI	1	2	2	4	5
	25,0 %	33,3 %	66,7 %	44,4 %	38,5 %

Messtechnik				
mittel	hoch	höchst	hoch/höchst	gesamt
4	4	4	8	12
25,0 %	0,0 %	25,0 %	12,5 %	16,7 %
1		1	1	2
1	1	1	2	3
2	3	2	5	7
50,0 %	75,0 %	50,0 %	62,5 %	58,3 %

Gebiet	Methoden/Apparate				
Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
Anzahl Dokumente	3	2	1	3	6
nur der Mensch	33,3 %	50,0 %	100,0 %	66,7 %	50,0 %
beide	1	1	1	2	3
nur die KI	2	1		1	3
	66,7 %	50,0 %	0,0 %	33,3 %	50,0 %

Verfahrenstechnik/Maschinenbau				
mittel	hoch	höchst	hoch/höchst	gesamt
4	5	1	6	10
50,0 %	40,0 %	100,0 %	50,0 %	50,0 %
2	2	1	3	5
1				1
1	3		3	4
25,0 %	60,0 %	0,0 %	50,0 %	40,0 %

Mittelwerte (über 20 Recherchen zu Erfindungsmeldungen):

Relevanz	mittel	hoch	höchst	hoch/höchst	gesamt
nur der Mensch	37,2 %	32,5 %	49,3 %	37,3 %	36,5 %
nur die KI	47,5 %	54,3 %	43,8 %	51,9 %	51,0 %

Fazit: Bei beiden Recherchetypen zeigt sich anhand der Einzelfälle kein klarer Trend hin zum Menschen oder hin zur KI; im Schnitt findet der Mensch bei beiden Recherchetypen mehr höchstrelevante Dokumente, die die KI nicht gefunden hat, als umgekehrt.