

Nachrichten aus dem Datenraum

Vernetzte Worte

Vermutlich jeder kennt das Problem. Sie interessieren sich für ein für Sie neues Thema und wollen sich zunächst einen Überblick verschaffen. Suchen Sie nach dem Thema, z.B. „Smart Farming“ im Web of Science, so finden Sie über 3.000 Publikationen. Viel zu viele, um alle davon zu lesen. Zunächst möchten Sie wissen, welche Teilthemen in diesen Publikationen behandelt werden - wie das Thema strukturiert ist. Eine Möglichkeit, diese Problematik mittels daten-analytischer Methoden zu adressieren sind Keyword-Analysen.

Im Web of Science werden zu vielen Publikationen sogenannte „Author Keywords“ angegeben. Das sind mehrere Worte oder Wortkombinationen, die den Inhalt einer Publikation umschreiben und von den Autoren der jeweiligen Publikationen stammen. Bei einer Publikation, die sich mit Drohnen in der Landwirtschaft beschäftigt, könnten Keywords wie z. B. „Artificial Intelligence“, „Smart Farming“, „Unmanned Aerial Vehicles“, „Drones“ und „Machine Learning“ auftauchen. Natürlich können nun einfache, statistische Verfahren genutzt werden, um herauszufinden, welches Keyword über die Zeit hinweg wie häufig vorkommt. Die Erfahrung zeigt jedoch, dass dies nicht genügt, um die Struktur des Themas zu erfassen.

Um dies zu erreichen, ist es sinnvoll, die Begriffe miteinander in Beziehung zu setzen. Ein simpler Ansatz führt zunächst zu einem Netzwerk: Wenn zwei Keywords wie z. B. „Unmanned Aerial Vehicles“ und „Machine Learning“ gemeinsam in einer Publikation auftauchen, dann beschreibt diese gemeinsame Nennung, die sogenannte Kookkurrenz. Deren Anzahl bestimmt nun wie stark zwei Keywords miteinander verbunden sind.

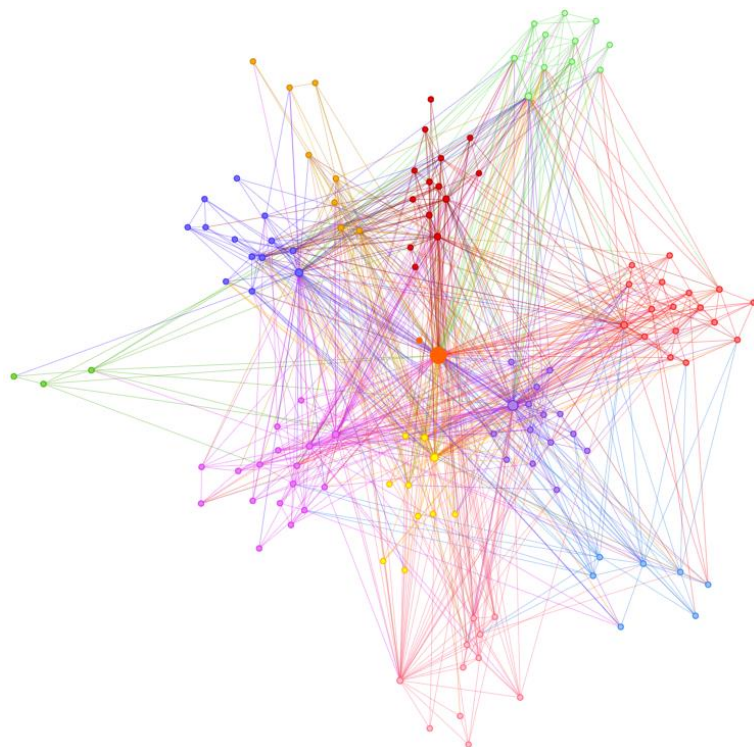


Abbildung 1: Darstellung eines Keyword-Netzwerkes, welches von Philipp Baaden aus der Gruppe Technology Foresight and University Hub (TFU, verlinkt) im Rahmen seiner Arbeit erstellt wurde. Die unterschiedlichen Farben der Knoten stellen die Zugehörigkeit zu unterschiedlichen Gruppen dar. Solche Darstellungen helfen dabei, die inhaltliche Struktur eines Themas zu verstehen.

Auf diese Weise entsteht ein Netzwerk. In diesem definieren die Keywords die Knoten und die Kookkurrenzen die Kanten. Bei mehreren tausend Publikationen kommt man meist entsprechend viele unterschiedliche Keywords und folgerichtig ein sehr großes Netzwerk. Es stellt sich also die Frage, wie dessen Komplexität so reduziert werden kann, dass die inhaltliche Struktur des Themas sichtbar wird.

Um dies zu erreichen, muss das Netzwerk in Gruppen von Knoten unterteilt werden, die besonders eng miteinander vernetzt sind – sprich in Gruppen von Keywords, die besonders häufig in Verbindung miteinander in wissenschaftlichen Publikationen auftauchen. Dies erfolgt mittels sogenannter Clustering Verfahren.

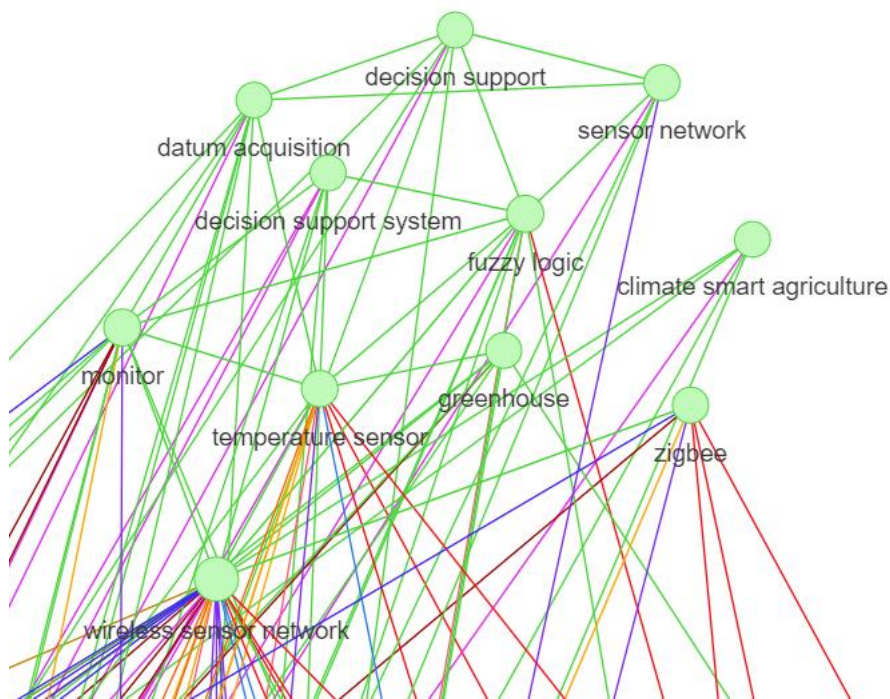


Abbildung 2: Deep-Dive in eine der Gruppen des Netzwerks. In dieser Gruppe geht es um den Einsatz von Wireless Sensor Networks für das Monitoring und die Entscheidungsunterstützung im Bereich der Landwirtschaft.

Die Analyse dieser Gruppen erlaubt es den Analysten, die inhaltliche Struktur eines Themas mit Hilfe der Keywords zu erfassen. Hierbei ist jedoch auch die Visualisierung entscheidend. Benötigt wird eine, die es dem Analysten sowohl erlaubt, einen Überblick über das Thema zu erarbeiten, als auch in die einzelnen Teilthemen einzutauchen und deren zeitliche Entwicklung zu erfassen. Daher sind interaktive Visualisierungen unerlässlich. Weitergehende Analysen wie das sogenannte Strategic Diagram sind von großem Nutzen. Dieses Diagramm versucht die einzelnen Gruppen auf Basis ihrer Rolle im Netzwerk zu kategorisieren. So lassen sich z. B. schnell grundlegende Themen von neueren, potentiell emergenten unterscheiden.

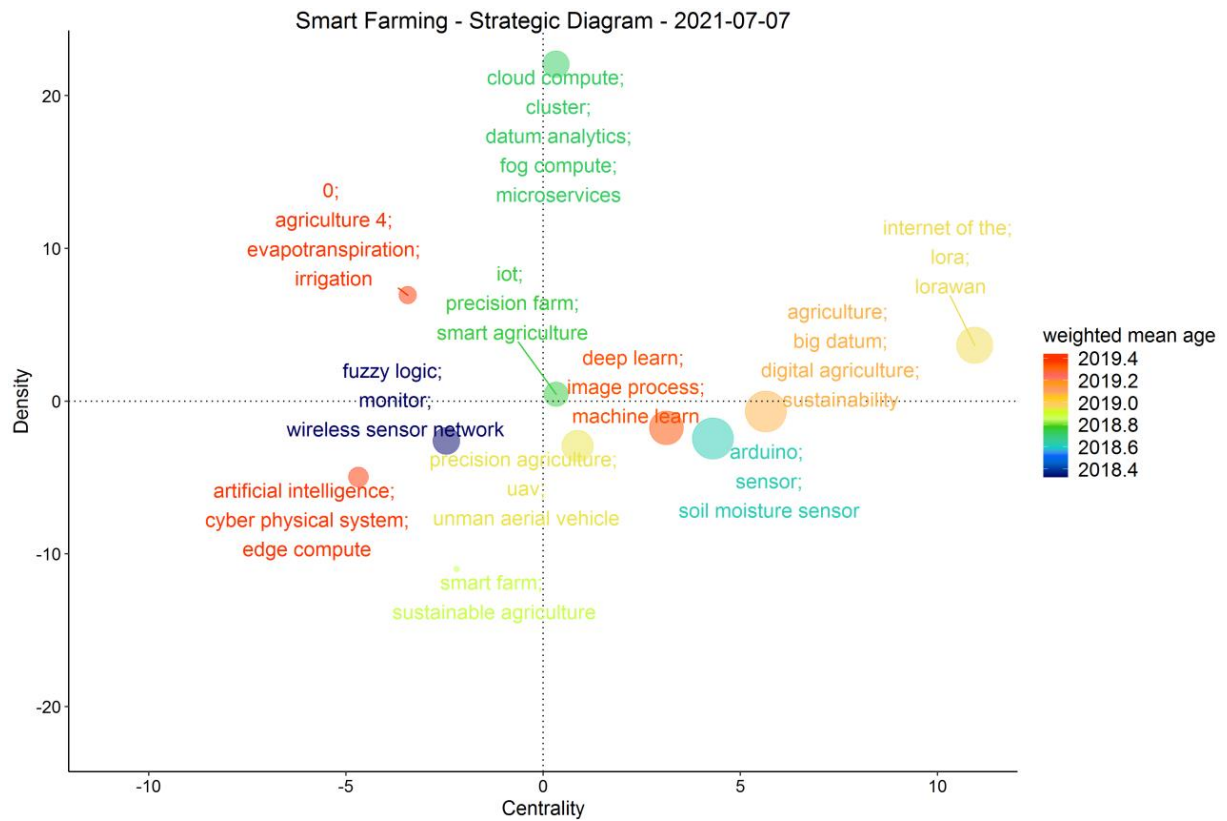


Abbildung 3: Strategic Diagram des Netzwerkes. Die Gruppen des Netzwerkes werden hinsichtlich ihrer Zentralität (wie stark sind andere Gruppen mit dieser Gruppe vernetzt) auf der x-Achse und ihrer Dichte (wie stark sind Knoten innerhalb der Gruppe vernetzt) auf der y-Achse verortet. Die Farbe beschreibt das mittlere Alter der Gruppe, gemessen anhand des mittleren Alters der jeweiligen Keywords. Emergente Themen sind in solchen Diagrammen meist im linken unteren Quadranten zu finden. Zentrale Themen meist im rechten, oberen.

Sie wollen wissen, wozu Netzwerke noch genutzt werden können? Dann bleiben Sie neugierig und datengetrieben.