https://www.researchgate.net/publication/221438417\_Linking\_a\_Domain-Specific\_Ontology\_to\_a\_General\_Ontology

Eine Ontologie von Konzepten in einem bestimmten Bereich, zusammen mit ihren Beziehungen und Eigenschaften, ist ein neues Medium zur Speicherung und Verbreitung von Fachwissen (Hsieh et al. 2010). Ontologien spiegeln eine spezifische Konzeptualisierung der Realität durch eine klare Definition von Konzepten (Begriffen) wider, die die Entitäten eines Fachgebiets und ihre Beziehungen darstellen. Aus einem linguistischen Blickwinkel definiert Sowa (2000: 492) eine Ontologie als "einen Katalog der Arten von Dingen, von denen angenommen wird, dass sie in einem Interessenbereich D existieren, und zwar aus der Sicht einer Person, die eine Sprache L verwendet, um über D zu sprechen". Eine Möglichkeit, die Elemente einer Ontologie anzureichern, besteht darin, sprachliche Informationen und Strukturen aufzunehmen (Buitelaar et al. 2009). In diesem Sinne ist eine mehrsprachige terminologische Wissensbasis eine wertvolle Wissensressource, da sie aus Zeichen in verschiedenen Sprachen besteht, die Konzepte bezeichnen, die mentalen Repräsentationen von Phänomenen der realen Welt entsprechen.

Hsieh. S.; Lin, H. T.; Chi, N. W.; Chou, K. W.; and Lin, K. Y.; 2010. Enabling the Development of Base Domain Ontology through Extraction of Knowledge from Engineering Domain- Handbooks. *Advanced Engineering Informatics*. doi:10.1016/j.aei.2010.08.004.

Sowa, J. F. 2000. *Knowledge Representation: Logical, Philosoph- ical and Computational Foundatio*ns. Pacific Grove, CA: Brooks/Cole.

Buitelaar, P.; Cimiano, P.; Haase, P.; and Sintek, M. 2009. To- wards Linguistically Grounded Ontologies. In *Proceedings of the 6th European Semantic Web Conference.* 111-125. Berlin: Springer.

Selenium und Crowler

Selenium ist ein verbreitetes Tool, das für Web Scraping und Crawling eingesetzt wird, um Websites automatisch zu durchsuchen und Daten zu extrahieren. Mit Selenium können Programme die Benutzerinteraktion mit einer Webseite simulieren, indem sie verschiedene Browser-Treiber verwenden, um auf der Seite zu navigieren, auf Links zu klicken, Formulare auszufüllen und Daten zu extrahieren. Dies macht es nützlich für das Web-Crawling, da Programme damit automatisch navigieren und Daten von Webseiten abrufen können, anstatt manuell durch jede Seite zu navigieren.

Die Inhalte der Website <https://www.br.de/nachrichten/suche?param=> sind dynamisch, d.h. sie können asynchron geladen oder dynamisch aktualisiert werden, was das Abrufen von Daten mit herkömmlichen Web Scraping-Techniken erschweren kann. Mit Selenium kann die Interaktion mit der Website automatisiert werden, einschließlich des Wartens auf das Laden von Inhalten vor dem Abrufen von Daten. Außerdem ist die Suchfunktion auf der Website interaktiv, d. h. das Programm muss Suchanfragen eingeben und auf Suchschaltflächen klicken, um Ergebnisse zu erhalten. Mit Selenium ist es möglich, diese Interaktion zu automatisieren, so dass die Suchanfragen und Ergebnisse programmgesteuert abgerufen werden können.

Um die Daten zu extrahieren, verwendet das Programm bestimmte Elemente auf der Webseite, wie z.B. CSS-Selektoren oder XPaths, um die gewünschten Informationen zu lokalisieren und automatisch zu extrahieren. Die gesammelten Daten werden in eine JSON-Datei gespeichert und anschließend automatisch geöffnet, um die Inhalte mit Hilfe eines HTML-Parsers zu extrahieren und mit dem entsprechenden Namen zu speichern. Der Korpus wird dann auf einzelne Dokumente aufgeteilt, was in der weiteren Verarbeitung hilfreich ist, um das semantische Netz logisch und nachvollziehbar aufzubauen.