

ID	Titel	Beschreibung	Bewertungskriterien	Ref.
CF1	Fokussiertes Crawling von HTTP-Seiten	Der fokussierte Crawler verwendet vordefinierte Seed-Listen, um auf Web-Ressourcen zuzugreifen, diese zu durchlaufen und bietet Filtermöglichkeiten für die Verfolgung von Links zu anderen Web-Ressourcen.	Manueller Abgleich von Meta-Informationen bei Datenportalen mit den gecrawlten Daten.	AK11
CF2	Zugriff über verschiedene Protokolle	Der Crawler kann auf Daten über verschiedene Standardprotokolle zugreifen, z. B. HTTPS und FTP	Testfälle für Crawler-Konnektoren; Verfügbarkeit von gecrawlten Daten von Seiten, die diese Protokolle verwenden.	AK11
CF3	Periodisches Crawling	Funktion zum Starten von periodischen und iterativen Crawling-Prozessen, die vom Crawling-Framework unterstützt wird, ist in der OPAL-Administration verfügbar.	Verfügbarkeit dieser Funktion mit Unit-/Integrationstest.	AK4
CF4	Fetching	Relevante Rohdaten, die vom Crawler gefunden wurden, werden extrahiert und gespeichert.	Präzisions-/Recall-Evaluierung mit manuell annotierten Goldstandard-Stichproben von Quellen.	AK11
CF5	Analyse	Halbstrukturierte, geladene Daten werden analysiert, um nur auf relevante Teile zuzugreifen.	Präzisions-/Recall-Auswertung mit manuell annotierten Stichproben	AK12
CF6	Extraktion	Rohdatenstrings werden aus umgebenden Strukturdaten (z. B. HTML-Tags) extrahiert.	Extrahierte Daten enthalten keine Strukturinformationen.	AK12
CF7	Semantische, maschinenlesbare Datenspeicherung	Gefundene Daten sollten in einem Datenformat gespeichert werden, das geeignet ist, die Semantik der extrahierten Informationen zu handhaben, um sie in einem nachfolgenden Integrationsschritt zu verknüpfen. (z.B. Turtle, Terse RDF Triple Language)	Validierung der gecrawlten Daten.	AK7, AK9
CF8	Speicherung von zeitbezogenen Daten	Die Crawler-Komponente reichert extrahierte Daten (siehe CF7) mit einem Zeitstempel an und speichert sie.	Unit-Test.	AK17, AK4
CF9	CKAN-API-Unterstützung	Der Crawler kann die CKAN-API für einen effizienteren Datenzugriff nutzen.	Integrationstestfall.	-