

Randnotizen des produktiv Unsicheren: Über die Dokumentation von Sackgassen

Kerstin Jung, Steffen Pielström, Patrick Helling

Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte, Institut für Deutsche Philologie, Universität Würzburg

Von Wegen des produktiv Unsicheren ...

... zu Gründen für den Verlust von Erkenntnis

- Neue Ansätze durch (Kombination von)
 - Daten
 - Methoden
 - Werkzeugen
 - ...
 - können zu Ergebnissen führen, die sich in Publikationen, Ressourcen und Dokumentation finden
 - oder in einer Sackgasse enden, wenn die Untersuchung ohne Ergebnis bleibt oder keine Ressource erzeugt werden kann.
 - Dabei ergibt sich bei beiden Alternativen oft ein Erkenntnisgewinn. Bei der Sackgasse z.B. in mehreren der von Gengnagel (2022) aufgeschlüsselten Problemfelder: Versagen von Technologien, menschliches, arbeitspraktisches sowie intellektuelles Versagen
- Im Falle einer Sackgasse kann die Erkenntnis oft gar nicht, oder nur implizit über weiterführende erfolgreiche Ansätze verfügbar gemacht werden.
 - Nach der Sackgasse muss die Arbeit meist schnell in eine andere Richtung fortgesetzt werden: Es bleibt keine Zeit eine ausführliche Dokumentation der Sackgasse anzufertigen.
 - Kein oder wenige Orte um diese Erkenntnis im Sinne von Negativresultaten oder basierend auf Präregistrierung zu publizieren.
 - Die Beschreibung einer Sackgasse ist nur dann nützlich, wenn sich zukünftige Vorhaben mit dem Vorgehen ausreichend vergleichen lassen
 - detaillierte Beschreibung der automatischen wie manuellen Verarbeitungsschritte
 - lässt sich die Sackgasse als daten-, werkzeug- und / oder methodengebunden identifizieren?

Kontinuierliche Prozessdokumentation

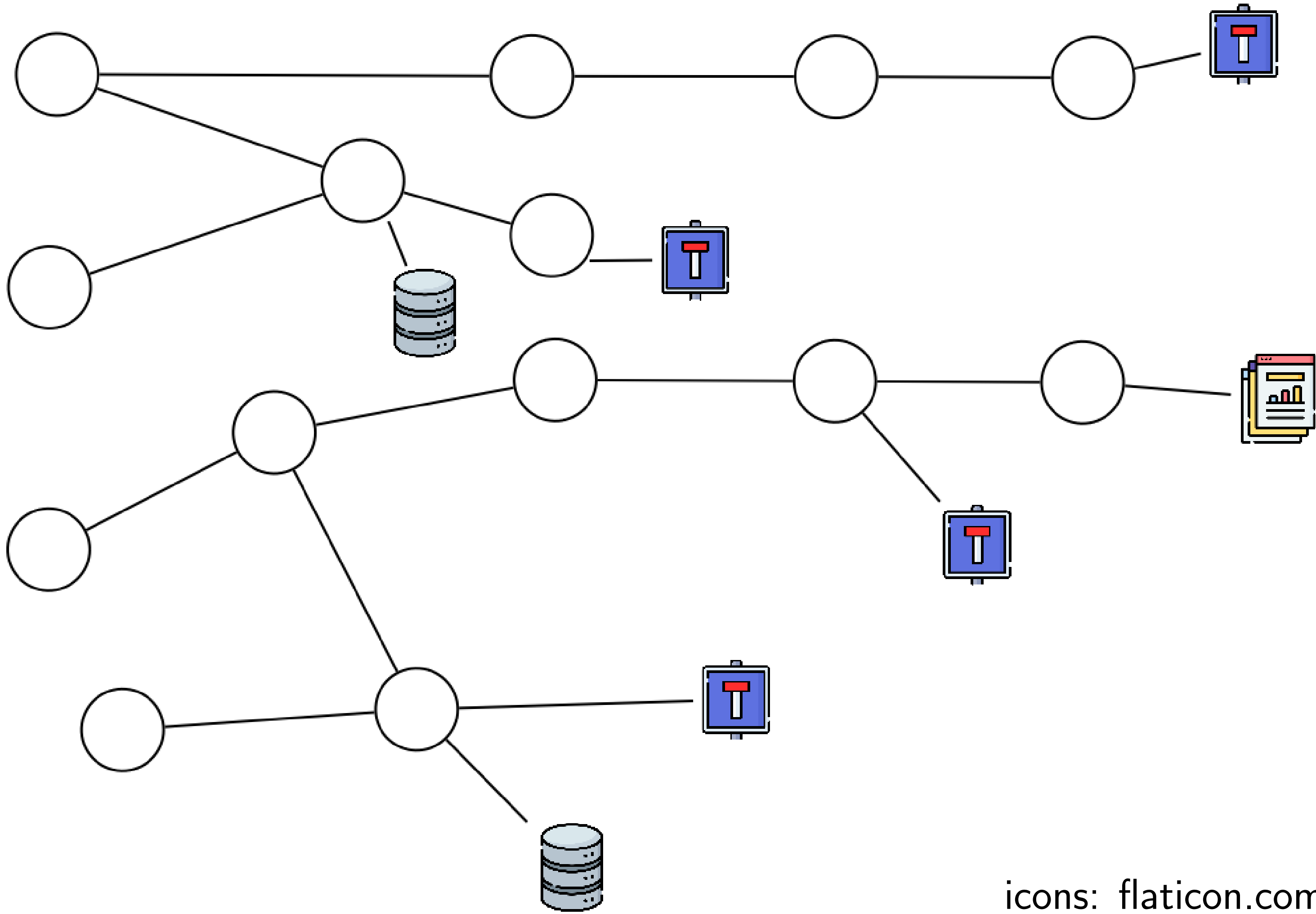
Prozessmetadaten um manuelle und automatische Schritte gleichermaßen zu erfassen

<input type="checkbox"/>	Auf Grundlage welcher Daten (+ Version) wird der Schritt durchgeführt?
<input type="checkbox"/>	Welche Operation wird auf den Ausgangsdaten durchgeführt? (Analyse, Selektion, Interpretation, ...; manuell, automatisch, semi-automatisch)
<input type="checkbox"/>	Wer oder was führt die Operation aus?
<input type="checkbox"/>	Person(en): Name / Kürzel / ID (Datenschutz bedenken)
<input type="checkbox"/>	Werkzeug(e): Name / ID, Version, Einstellung und Parameter, ggf. Betriebssystem / Hardware
<input type="checkbox"/>	genutzte Komponenten (Annotationsrichtlinien, Lexikon, Modell, etc.): Name / ID, Version
<input type="checkbox"/>	ggf. Zusammenhänge zwischen Ressourcen (Korpus + Version auf dem das Modell trainiert wurde, ...)
<input type="checkbox"/>	ggf. neu entstandene Daten (Subkorpus, Annotationsebene, ...): Name / ID, Version

- + Dokumentation der Wege, nicht bezogen auf ein bestimmtes Resultat
- + Umfasst Dokumentation von Sackgassen, da sie beim Erkennen der Sackgasse bereits vorliegt
- + Detaillierte Dokumentation erhöht die Vergleichbarkeit von Prozessen → Erkenntnisgewinn aus Sackgassen wird vereinfacht

Eine Galerie des gut dokumentierten Scheiterns? – Fragen an die Community

- Würden wir eine Datenbank der Sackgassen vor Beginn unserer Vorhaben durchsuchen?
- Wären wir bereit mit unseren Prozessbeschreibungen zu einer solchen Datenbank beizutragen?
- Könnten wir uns im Fachbereich auf ein (oder mehrere abbildbare) Format(e) zur Prozessdokumentation einigen?
- Ist es realistisch, dass wir unsere Arbeitsabläufe forschungsbegleitend (und ggf. teil-automatisch) mit entsprechenden Prozessmetadaten versehen?
- Können wir die forschungsbegleitende Prozessdokumentation in Schulungen und ggf. Studiengänge aufnehmen?
- Wie schätzen wir die Vergleichbarkeit einzelner Ketten von Verarbeitungsschritten ein?



icons: flaticon.com