

Personal Knowledge Management Systems for Capturing AI-generated Content

Clemens H. Cap

October 2026

1 Introduction

For the last 20+ years we have seen an increased trend to deliver information through the Web. Content presentation is done in various ways via static or dynamic web pages, on blogs, forums, content management systems or wikis, free or for closed user groups or behind paywalls, as part of highly dynamic services such as timelines, chat or AI systems, in widely different original formats from HTML to Markdown to image or database formats. Information is often kept in proprietary, closed or access regulated solutions.

Before the age of digitization, the user had to take the burden of putting down manual notes or even making physical copies to ensure continuous access to consumed information. During the onset of digitization most information was freely accessible and in plain text format. In our age of more advanced digitization the range of possibilities of storing information in personal knowledge management systems has become larger and all kinds of formats can be employed. There are numerous systems which aim at storing notes for the user. However, most systems have **inherent pain points**.

For example: (1) HTML pages can be saved, but mostly all assets of the HTML pages are saved, including css and javascript. (2) Web clippers work for Evernote and OneNote, but these systems are proprietary and data formats, storage and workflow are not fully under the control of the user. (3) Printing a page as PDF or snapshotting it as BMP stores data in a too complex and not so easy editable format, moreover information is stored in the originally presented look-and-feel and not in a simple and uniform format suitable for knowledge management. (4) Apple music playlists can be shared via Apple's infrastructure, but exporting it into TXT or CSV format requires external tools such as exportify or soundiiz or needs additional quirks to be able to access all metadata. (5) AI systems such as ChatGPT provide possibilities to share chats, but they completely remain under the control of the bot-

For an overview on concepts, see <https://chatgpt.com/share/68dba172-f5c0-800f-a3ef-6a9b48d449e0>

As a result the individual users lose access to information they gathered and managing knowledge in a manner where structure, search, access and formatting is under the control of the user becomes increasingly tedious – although digital systems are in principle capable of offering flexible suitable user and data interfaces, APIs and metadata formats.

2 Research Items

1. What concepts might a user need to manage information collected by AI bots or similar systems?

Design a concept for provider-independent, system-independent, user-centered knowledge management of information collected by the user in the browser. The concept should consist of cognitive elements, key principles, methods and workflows for collecting, tagging, attaching metadata. summarizing, annotating, (re)formatting, updating and more elements which a user might require during the life cycle of the collected information.

2. How can these concepts be mapped to data structures and APIs.

Task: Turn the concept into an interface (API description which should be adjusted to a client-server kind of architecture). As client assume a standard browser extension, as server assume a Mediawiki installation.

3. Task: Build a proof-of-concept for this architecture in form of a Chrome extension. Information sources: ChatGPT, Perplexity, Claude, Wikipedia, generic web page. Information sink: Generic Mediawiki installation.

The use of AI methods for writing this thesis and generating the proof of concept is encouraged. Token costs can be subsidized by the Chair.

Requirements: The successful candidates for this topic have a good knowledge of Javascript and feel comfortable with vibe-coding concepts.