From Semantic Web to Wisdom Web: A Retrospective on the Journey to 2043

Gary P. Tchat¹, Amin Anjomshoaa²,³, Daniil Dobriy², Fajar J. Ekaputra², Elmar Kiesling², Axel Polleres² and Marta Sabou²

¹Council on the Equal Distribution of AI (CEDAI), Brussels 1000, Belgium
²Institute of Computer Wisdom, WU Wien Vienna 1020, Austria
³Complexity Science Hub Vienna Vienna 1080, Austria

Abstract

Forty years ago, the vision of a “Semantic Web” (SW) started out from prerequisites that today are difficult to imagine – it was conceived as an extension of what was then the World Wide Web (WWW) and was supposed to enable computers to understand and process information in a more human-like way [1]. The original aim was to accomplish this through the representation of information using structured formats, in combination with linking, and sharing of semantic resource descriptions: the term “semantic” was meant to express conveyance of meaning to...

Keywords
Wisdom Web, Large General Models, Universal Language

DISCLAIMER: This paper is a work of fiction, written in 2023 and describing research that may be carried out in and until 2043. For this reason, it includes citations to papers produced in the period 2024-2043, which have not been published (yet); all citations prior to 2024 refer instead to papers already in the literature. Any reference or resemblance to actual events or people or businesses, past present or future, is entirely incidental and the product of the authors’ imagination. Even the imaginary 2043 keynote speaker and first author, who started its PhD in the early 2020’s, is fictitious.

1. Introduction

Forty years ago, the vision of a “Semantic Web” (SW) started out from prerequisites that today are difficult to imagine – it was conceived as an extension of what was then the World Wide Web (WWW) and was supposed to enable computers to understand and process information in a more human-like way [1]. The original aim was to accomplish this through the representation of information using structured formats, in combination with linking, and sharing of semantic resource descriptions: the term “semantic” was meant to express conveyance of meaning to...
marches through symbolic, compact and unambiguous forms in between computational – back then – discrete numerical representations and human-written/spoken language. This original idea of representing knowledge explicitly to make it more “machine-processable” can be interpreted as a way to cope with the limited computational capabilities characteristic for the time, but also as a way to open that knowledge up for large-scale symbolic computation.

In today’s Wisdom Web (WW), mass-scale machine interpretation and agent interaction have been accomplished, although through very different means than originally envisioned – including the recent emergence of a Universal Language and the evolution of the Knowledge Continuum that no longer separates between structured and unstructured representations. It is interesting to note that already in the early days, doubts were raised about the continuous scale of the very notion of “symbols” [2], as well as on the strict distinction between discrete identifiers and meaning [3] in this past Semantic Web (SW).

Other traces of the original SW vision are also reflected in the WW, including decentralization through fluid transition between localization and global integration in shared cognition spaces, and agent-mediated collaboration – even though these “agents” interact differently and much more autonomously than envisioned. New challenges have emerged – partially widening and partially narrowing gaps between the knowledge bases and reasoning and communication capabilities of humans and Artificial Intelligence (AI) agents we see today.

Indeed, research around wisdom maintenance has evolved tremendously to drive and shape these developments over the past four decades and shifted its focus repeatedly – most recently from knowledge to wisdom, so we may well call the current WW community a natural evolution of this past SW community. This paper aims to lay out a timeline of these developments structured into three major periods (cf. 1): Section 2 covers early milestones from the original SW vision to the rise of Knowledge Graphs (KGs); Section 3 discusses the revolutionary developments of the 2020s and Section 4 summarizes more recent events up to 2043, including the transitions that shaped today’s Wisdom Web. We close in Section 5 with some questions to the past.

Figure 1: The Evolution of the Wisdom Web
2. Early Milestones (2001-2021)\textsuperscript{1}

The original \textit{Semantic Web Vision} was introduced in a seminal article in 2001 and conceived the SW as an extension of the WWW that allows data to be shared and reused across applications, platforms, and organizations [1].\textsuperscript{2} It aimed to create a “Web of Data” that can be easily understood and processed by “intelligent agents”, enabling them to perform tasks such as searching for information, completing transactions, and making decisions on behalf of humans.

In the mid 2000s, the field increasingly recognized that the SW vision had not been delivered, but researchers argued that the development of strong standards would enable large-scale, agent-based mediation in the future [9]. Even though the development of standards proceeded to create foundations for the representation, querying, and sharing of structured data across the Web, standards for other parts of the envisioned SW stack remained incomplete or saw limited adoption. This included ideas for (i) a \textit{unifying logic} on top of conceptual symbolic representations (often termed Description Logics, which were, however, limited in expressivity by boundaries of computational tractability) and \textit{rules}, (ii) encryption mechanisms to protect or signing data, providing \textit{proof}, and eventually (iii) \textit{trust}, as well as (iv) bespoke user interfaces and applications. In hindsight, we may argue that the low adoption was on the one hand due to the representational gap of the symbolic representations and on the other hand due to the limitations of known reasoning methods at the time, as compared to actual, actionable wisdom.

\textbf{Linked Data.} On the other hand, the community back then already recognized that a second missing piece needed was connectivity between data (and agents); gaining traction in the early 2010s, the second wave of research was trying to close this gap with approaches centered around the idea of Linked Data [10], which, building on the ideas of the original WWW, emphasized the practice of connecting data through common, unique identifiers, enabling efficient data integration and knowledge discovery. Despite impressive initial growth, however, the concept failed to gain widespread adoption in industry and mainstream Web applications. Proposed explanations in the literature include questionable data quality [11] and the technical limitation of relying on discrete, uniform identifiers based on binary computer representations. These limitations made it note only unfeasible to evaluate even simple queries on the available Linked Data Cloud within reasonable time boundaries [12], but also implied that query results for any dynamically represented data were provably incomplete [13].

\textbf{Knowledge Graphs} The third wave of research represented a shift in focus from decentralized data publishing on the Web towards the construction of increasingly sophisticated, large-scale, and more loosely interconnected KGs [14]. Early examples of open and commercial KGs – including DBpedia [15], YAGO [16], and Google’s KG – were typically constructed from large text corpora such as Wikipedia\textsuperscript{3}, whereas Wikidata [17] represented an early initiative to collaboratively


\textsuperscript{2}Although this article is commonly cited as the inception of the SW research community, the history of key ideas goes back much further – cf. [8] for a review of early historical events.

\textsuperscript{3}A community-driven project that was initially edited by human collaborators.
build a large-scale KG through volunteer contributions from scratch. Interestingly, Wikidata, which already inherently incorporated the idea of collaborative knowledge evolution, has become absorbed into today’s Wisdom Web and is largely being evolved through knowledge reflection by Large General Models (LGMs), which are also based on collaborative, interactive knowledge aggregation, but no longer restricted to mostly human contributors only.

Overall, the first two decades of research were characterized by alternating proclamations of ambitious visions, substantial progress in research, recognition of (often non-technical) factors that limited the viability of the visions, and ultimately adoption, adaptation and re-invention of key ideas in industry.

3. Advancements in the 2020s

The third decade of the 21st century saw rapid advancements in the SW research community characterized by a number of broader developments that shaped the evolution of the field. Most notably, although trust and explainability had been recognized as crucial issues in AI research early on [18], the diffusion of Large Language Models (LLMs) in the mid 2020s increasingly transformed the issue from an academic concern to a contentious and impactful real-world issue. Paired with a general tendency towards secrecy in the late 2020s, this led to an erosion of trust that accelerated as fundamental limitations of LLMs w.r.t. their ability to learn abstract concepts and representations became increasingly evident and as reinforcement learning-based guardrails proved ineffective in the face of adversarial techniques. This led to an increasing recognition that concerns around transparency, reliability, and trust were inherently difficult to overcome, triggering a second wave of responsible AI research that strongly influenced developments in the SW research community.

(Deep) Neuro-symbolic convergence. Large KGs that had become available at the end of the 2010s highlighted the scalability limitations of classical reasoning mechanisms and, as an alternative to those, prompted the use of sub-symbolic AI techniques for processing and evolving them. Conversely, despite the early successes of (Deep) Neural Networks to solve specialised tasks, the need for more complex knowledge representation emerged. Such developments led to the idea of combining the strengths of techniques from both sub-fields of AI and the quest to build (deep) Neuro-symbolic Systems (NeSys) – a vision that was labeled "the third AI summer" [19]. In the SW community [20], this was reflected in an increased interest in combining SW and Machine Learning (ML) components [21] into novel systems combining the representation, reasoning and explainability features of symbolic systems with the ability to learn from large amounts of data.

Research focused on both creating and understanding the characteristics and optimal design of NeSys. A pattern language for representing NeSys greatly facilitated this endeavour. Within a few years, the field converged to generally approved NeSys design patterns, a cornerstone of the maturing field of AI system engineering. This made it possible for AI engineers to describe and even automatically generate complex NeSys, which also revolutionized the construction of KGs.

However, as they became increasingly complex, NeSys encountered a challenge in seamlessly integrating the various representation and inference mechanisms of their components within
the same system architecture – a gap addressed by the emergence of polyglot representations towards the end of the 2020s: bespoke neural models and architectures, designed based on said NeSys architectures could be be deployed to interlink and translate between neural numeric representations and symbolic knowledge bases, and provided the first robust solution to not only seamlessly integrate axiomatic rules and constraints with neural architectures, but also to ringfence the detrimental artifacts stemming from self-feeding purely neural language models. Yet, hardly remembered by today’s standards, the KG systems at the end of the 2020s were still operating on separate traditional numeric and symbolic presentations, in particular (i) manually human-crafted symbolic constraints safe-guarding neural inference, and (ii) the polyglot representations paradigm that was mainly based on parallel symbolic or neural representations that were hard-wired through interpreting models with fixed neural architectures. So, while the integration of compact symbolic presentations could improve robustness, as well as scalability in use cases such as multi-modal conversational agents, the underlying computing technology was still limiting in terms of the maximum NISQ [22] integration of quantum computing hardware.

### Knowledge Graph Construction and Evolution

In the mid 2020s, rapid advancements made it possible to mass-construct and update domain-specific KGs robustly at scale. Whereas in the early 2020s, KG construction was still solved by hand-crafting pipelines that carefully orchestrated a variety of tasks [23], research focus in the second half of the 2020s increasingly shifted towards direct multi-modal KG construction fusing deep learning approaches with generative KG construction techniques. This was enabled through polyglot representations that allowed human actors to interact with knowledge construction without the need to explicitly formulate constraints symbolically. Driven by the rise of LLMs, another large stream of research emerged on methods that cast the knowledge contained within these models into explicit KGs representations [24]. Initially developed to construct KGs from unstructured text, these methods increasingly became used to explicate and materialize, as well as to (cross-)validate LLMs, e.g., by facilitating conversations between model instances and negotiating consensus [25]. In the late 2020s, concerns about the reliability of (Deep) Learning Models also motivated research on rule- and constraint-based engineering techniques for LGM [26]. In this context, KGs proved to be useful as a canvas for introspection and tremendously improved the self-reflective capabilities of LGMs.

### Polyglot representations and inference

The 2020s saw a growing variety of graph-oriented data models, persistence systems, and query languages as well as a shift in written language as such to cater to increasingly diverse storage and reasoning needs. This led to a proliferation of architectures based on the above-mentioned concept of “polyglot representation” [27] developed to enable fluid transitions between various storage and querying paradigms as well as between emerging axiomatic knowledge representations that were no longer only restricted to steer neural computation, but also being generated from neural models as part of the explanation process. Polyglot representation provided uniform access across heterogeneous graph models, storage technologies, semantic and latent representations, and query languages, giving rise to a shared cognitive environment for graph-centric AI applications. This resulted in a large stream of research on how to manage replication issues and dynamic transition and led to the development of
RXF [28] as a protocol to facilitate coordination between decentralized polyglot knowledge nodes. Although this enabled fluid alignment of structured representations with natural language as well as advances on polyglot inference mechanisms that combined neural and symbolic techniques, it also increased data and knowledge fragmentation and did not ultimately solve interoperability challenges.

4. Advancements in the 2030s and early 2040s

Following global societal disruptions in the late 2020s, which were linked by many researchers to the large-scale transformation of virtually all economic sectors through AI [29], the early 2030s were characterized by the emergence of a more cooperative culture. Following the initial friction created by the progress made in intelligence augmentation and creativity-enhancement, the early 2030s finally saw the emergence of strong Human-AI (HAI) collaboration. Consequently, continuous collaboration became the dominant paradigm for most tasks that still involved humans, giving renewed urgency to fundamental questions about the nature and role of knowledge and reasoning. Driven by these broader developments, SW research was redefined dramatically in the wake of the emergence of the Wisdom Web (WW) and drove the convergence towards unified entangled description representations, unified languages, and enabling personal AI companions.

**Wisdom Web**

Emerging in the 2030s, the Wisdom Web (WW) has been a pivotal development in computer-assisted wisdom research. It provides a comprehensive knowledge infrastructure that connects various data sources, LGMs, and personal AI companions to facilitate seamless Human-AI interaction. While the introduction of semantics in data communications marked a significant milestone in the evolution of information theory [30], the Wisdom Web can now explicate and capture non-deterministic, fuzzy, and tacit knowledge [31]. Wisdom Web consists of multiple layers, including a data layer for integrating the continuum of structured and unstructured data from diverse sources, as well as the Universal Description Framework (UDF) knowledge representation layer that has in the meanwhile almost entirely replaced the traditional polyglot representations, and a reasoning layer that seamlessly supports both neural and symbolic approaches [32]. WW’s decentralized architecture ensures that knowledge is distributed across various nodes, enhancing scalability and resilience [33]. To this end, distributed ledgers, which are an integral part of the WW, facilitate trustworthy and transparent communication between AI systems. This structure enables the WW to serve as a nexus for vast amounts of knowledge, promoting a more efficient and accessible cognitive environment for intelligent AI agents and human users.

Remarkably, the nowadays common UDF [34] standard can be seen as a natural evolution from the original RDF and knowledge graphs representation, enabled by the recent 2nd quantum wave via the intermediate polyglot representations combined with RFX. Only after the NISQ limitations that restricted quantum computers to small representational spaces or very limited operational depth had been overcome through the modern CISQ architectures was it possible to design adequate representations that could capture KGs in terms of their edge space and their interactive evolution in one unified model. The QHedeges (Quantum-Hyper-Edges) model was a fundamental breakthrough for this and enabled UDF as a foundation of the WW. The instant and
entangled representations enabled by UDF both replaced separate polyglot representations and the necessity to explicitly model "links": the Wisdom Web vision entirely can rely on entangled links at super-global scale.

**Personal AI Companions** Early SW-enabled techniques to support HAI blended human expertise with machine intelligence for improved accuracy and efficiency already in the late 2020s. As part of this development, human language increasingly became the dominant interface between all agents (human and AI) in most HAI scenarios, giving way to human-centric collaborative models. As a further development, powered by UDF representations, the Wisdom Web has led to unprecedented advancements in the capabilities and performance of AI companions who utilize Universal Language (UL) to interact seamlessly with humans and other intelligent agents in diverse and complex scenarios. We should emphasize though, as private AI companions are increasingly perceived as a core part of the modern family [35] and brand AI companions are transforming the service economy [36], many ethical questions still remain unanswered: despite far-reaching consensus on the categorical equivalence of human and advance AI cognition [37], AI rights are still a fringe topic in the ultimately human-centric AI policy [38].

**Universal language** After natural language had already become the standard interface between and among humans and AI agents, it became increasingly apparent that LGMs started to evolve their own, much more efficient communication patterns and languages. Consequently, research no longer focused on human-designed polyglot representations, but on naturally evolved malleable languages that were exceedingly difficult for humans to analyze. Computational linguists were surprised by how fast these languages evolved, raising considerable ethical concerns. Not only were the decisions and actions of individual AI agents still not explainable, but humans were no longer able to understand their interactions, even though they had already influenced the evolution of human-to-human communication significantly when compared to 20 years prior. As mentioned in the abstract, when reading the present paper, whose language may nowadays seem archaic, we can observe how strongly the increased human-to-machine and machine-to-machine interaction shaping discourse and knowledge evolution have shaped even our written language in the meanwhile.

5. Conclusions

As a summary, we play a small thought game, hoping our work is being “sent back to the past”: let us attempt to formulate the big advances of the past decades as research questions that people should have asked 20 years ago in order to arrive where we are now:

- Which representations can be used to bridge in between embeddings and symbolic representations of Knowledge Graphs?
• How will conversational AI agents interacting not only with humans but also amongst each other influence written and spoken language as well as the Semantic Web visions original idea of Personal (AI) agents?

• Apart from what we now call knowledge representation, how can computational artifacts and models that also track interactions between agents and communities capture and represent (organisational) “wisdom”, commonly referred to as “tacit knowledge” [31] nowadays?

• Can quantum computing and entanglement be used for knowledge representation and decentralized processing?

Only future will tell.

References

Human-AI Collaboration 12 (2040) 15–34.


[34] R. Denenberg, J. Kunze, Universal Description, Discovery, and Integration (UDDI) - Primer, W2C Recommendation, Wisdom Wide Consortium (W2C), 2037. URL: https://www.w2.org/TR/2037/UDF-primer/.


