Question-answering system for combustion kinetics

Laura Pascazio[1], Dan Tran[1], Simon Rihm[1,2], Jiaru Bai[2], Jethro Akroyd[1,2,3], Sebastian Mosbach[1,2,3], Markus Kraft[1,2,3,4,5]

released: December 08, 2023

[1] CARES
Cambridge Centre for Advanced
Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[2] Department of Chemical Engineering
and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[3] CMCL Innovations
Sheraton House
Cambridge, CB3 0AX
United Kingdom

[4] School of Chemical
and Biomedical Engineering
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

[5] The Alan Turing Institute
2QR, John Dodson House
96 Euston Road
London, NW1 2DB
United Kingdom

Preprint No. 315

UNIVERSITY OF
CAMBRIDGE

Abstract

In this paper, we introduce for the first time a natural language question-answering (QA) system specifically designed for the field of combustion kinetics. This system marks a significant step towards achieving the PrIMe vision as outlined by Frenklach in 2007, offering a user-friendly interface that allows researchers and practitioners to easily access and query information about chemical mechanisms. This QA system is a key component of "The World Avatar" (TWA), a dynamic framework built upon semantic web technologies. TWA is characterized by its layered structure, which includes a knowledge graph, software agents, and real-world data integration. These layers collectively create a comprehensive unified system for managing and analyzing complex chemical data from various domains. We detail the enhancements made to TWA's ontologies (OntoSpecies, OntoKin, and OntoCompChem) to meet specific challenges in chemical kinetics and improve their representation accuracy. By focusing on data provenance and interoperability, our approach ensures transparent and reliable data management that adheres to the FAIR principles, which is vital for precise information retrieval and analysis. The role of software agents in populating these ontologies is highlighted, showcasing how they transform raw data into meaningful structured knowledge and generate new insights within the TWA ecosystem. Additionally, the semantic web technologies' interoperability feature facilitates data integration and exchange across different platforms and tools, making the data machine-actionable. We then demonstrate the QA system's capabilities in answering questions related to four H2/O2 reaction mechanisms taken from the literature as a proof of concept. Lastly, we discuss the future directions of the TWA framework, which include not only future extensions of the QA system but also the integration of external tool to automate tasks such as generation of kinetic mechanism, further expanding TWA's functionality and application in the field of chemical kinetics.
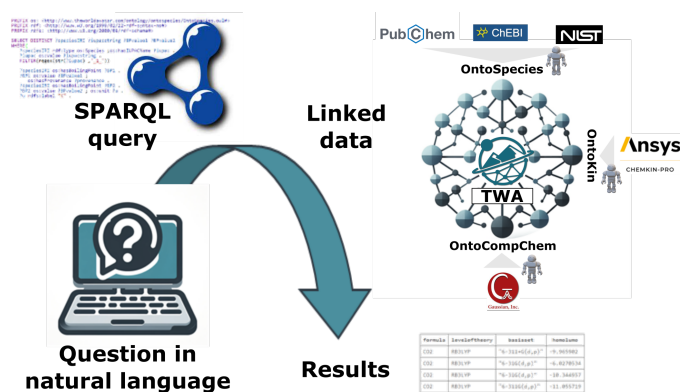
Highlights

- A natural language question-answering system has been specifically developed for combustion kinetics.

- The system builds upon a dynamic knowledge graph based on semantic web technologies offering a unified approach to handling complex chemical data.

- Our approach combines a user-friendly access with rigorous data management that makes combustion kinetics data more accessible and machine actionable.

Contents

1 Introduction

Combustion is an exothermic process that fundamentally powers our modern world, ranging from internal combustion engines to industrial furnaces [18]. As we progress towards the United Nations' Sustainable Development Goals (SDGs), the precise prediction and control of the behavior of reacting mixtures during combustion reactions become increasingly important. The key to success is accurate kinetic modeling as it encapsulates our understanding of the underlying chemical reactions and their rates.

Modeling combustion kinetics accurately is not an easy task due to the inherent complexity and multifaceted nature of combustion processes. Even in the combustion of relatively simple hydrocarbons [4], a vast number of chemical species and reactions emerge, resulting in highly non-linear behavior covering a broad range of time and length scales – from nanoseconds to milliseconds and from molecular to macroscopic levels. The sensitivity of reaction rates to temperature, described by Arrhenius equations, means that minimal fluctuations in temperature can lead to significantly different outcomes. The real-world interaction between turbulence and chemistry adds an extra layer of complexity, making the task even more challenging.

Compounding these challenges is the variability in the reaction mechanisms developed by different research groups. Depending on focus, methodologies, and available experimental data, diverse mechanisms can emerge for the same fuel [2]. While all aim to capture the combustion phenomenon, inconsistencies may arise when comparing these mechanisms, especially for reduced mechanisms extracted for practical applications [23].

This fragmented landscape necessitates a comprehensive system capable of efficiently managing extensive experimental data and kinetic mechanisms. This is crucial for detecting inconsistencies across mechanisms and predicting reaction rates for missing reactions. In 2007, Frenklach [9] introduced the Process Informatics Model (PrIMe) initiative, a data-centric approach to developing predictive models for complex chemical systems. It envisions an intelligent question-answering (QA) system to assist researchers during this process. One key feature of PrIMe lies in its method of organizing scientific data by merging diverse sources based on both data semantics and provenance. The initiative strives to maintain a complete and up-to-date set of data in combustion chemistry for rapid mechanism bench-marking and analysis. This data curation model holds great potential to foster democratic decision-making and consensus-building within the combustion community.

A significant hurdle in realizing this vision is the isolated nature of existing data, mechanisms, and software tools [20]. This not only impedes interoperability between the outcomes produced from different groups but also results in their unavailability once projects conclude or servers are shut down (e.g., CMCS [25]). Therefore, we aim to address this challenge by creating a collaborative framework following an open-access philosophy and a distributed web architecture. This design allows researchers to contribute their data and mechanisms to the ecosystem while maintaining full control and leveraging any preferred data source.

In this paper, we introduce a natural language question-answering system specifically designed for combustion kinetics that is aligned with and inspired by the PrIMe vision. This system is build upon a dynamic knowledge graph based on semantic web technolo-

gies as part of "The World Avatar", which provide an open-access unified framework for managing a wide array of complex chemical data from diverse chemical domains. Emphasizing data provenance and interoperability, our approach guarantees transparent and reliable data management, crucial for accurate information retrieval and analysis. Simultaneously, the interoperability aspect of semantic web technologies allows integration and exchange of data across diverse platforms and tools, adhering to the recommendations from Cavallotti [7] to streamline the integration of automated software. This integration that combines user-friendly access with rigorous data management, represents a significant advancement in making combustion kinetics data not only more accessible but also machine actionable. In the following sections, we will explore the design, capabilities, and potential applications of this system in combustion research.

2 Background

2.1 Challenges

The field of combustion kinetics has witnessed significant progress, especially in developing complex models for the pyrolysis and oxidation of a broad spectrum of hydrocarbons. Despite these advancements, the discipline confronts formidable challenges in formulating predictive kinetic models that maintain accuracy across a wide range of conditions [7, 12, 13]. particularly as the complexity of reactants escalates. These models must not only accurately replicate extensive sets of experimental data but also incorporate well-defined uncertainty bounds in their predictions [26]. This task is rendered arduous by the overwhelming volume of data and the intricate nature of accurately modeling chemical reactions within combustion processes.

A critical issue is the disorganized state of the collection, evaluation, and selection of chemical reaction models, leading to conflicting information and a disjointed understanding of kinetic models. This issue highlights the need for improved coherence and reliability in kinetic modeling, particularly through provenance tracking to ensure consistency, eliminate duplicates (e.g., permutations of the same reaction), and adhere to the FAIR principles – Findable, Accessible, Interoperable, and Reusable [1]. This cannot be achieved without undergoing a process of "transformation" of these disparate data sets into functional models, a process complicated by escalating data volumes. This situation underscores the importance of evolving databases from mere repositories into dynamic tools for knowledge creation, addressing the metadata challenge concerning underlying assumptions, parameter origins and uncertainties, reasons for inclusion/exclusion of reactions [12].

Additionally, the current landscape in chemical engineering, where many practitioners lack expertise in computing rate coefficients from first principles or using modern kinetic modeling software and machine learning techniques, calls for a robust computational framework that accurately determines thermochemical and transport parameters and rate constants for a wide array of chemical species and reactions.

2.2 Available solutions

In response to these challenges, various automated tools like RMG (Reaction Mechanism Generator) [10] and Genesys [33] are being developed to generate kinetic mechanisms, employing methods ranging from pre-defined reaction pathways to unbiased approaches. These tools, starting with a core set of species, expand the mechanism until additional species minimally impact reactivity. While these tools are advancing towards the goal of fully predictive kinetic models, gaps remain, particularly in evaluating thermochemistry and rate constants, highlighting the need for enhanced automation in these areas. A combination of theoretical and experimental methodologies can be used to enhance the predictive capabilities of these mechanisms [21]. This approach, coupled with automated procedures for the investigation of unknown reaction pathways, is critical for discovering new reactive pathways and achieving a comprehensive understanding of combustion kinetics. Additionally, tools like RMG are occasionally failing to select the most relevant species or reactions. This issue stems from RMG's reliance on reaction templates, which may not always cover all the necessary reactions critical to a particular system and to improve the tool's accuracy it is essential to expand the range of reaction templates within RMG [10]. The need to overcome this issue with automatic tools is subject of ongoing research.

Other essential tools that deal with different aspects of gas-phase kinetic modelling are also available – e.g., Chemkin [16], Cantera [11], ReSpecTh [34], PrIMe [9]. The challenges in working with these are mostly related to interoperability and accessibility. Moreover, duplicate information in different formats or databases is hard to spot [14].

In order to increase ease of access, QA systems can significantly lower entry barriers. The current wave of large language models illustrates this impressively in many areas. However, general NLP tools that are available to the public are monoliths in nature and lack explainability and domain knowledge (e.g., ChatGPT). Going forward, computational chemistry needs a "peer-to-data-driven operation with respect to molecular science to act autonomously on arguments and questions raised in natural language by the operator" [32].

2.3 Semantic Web for representing chemical kinetics

The Semantic Web offers an innovative, graph-oriented approach for effectively representing the intricate domain of chemical kinetics. It structures chemical information into a graph, aptly capturing the complex relationships and properties inherent in chemical processes. This approach is particularly well-suited for chemical kinetics, where many tools are graph-based [14, 30], as it naturally aligns with the graph-like nature of chemical interactions, depicted in Fig. 1.

In this framework, reaction mechanisms are portrayed as networks of interconnected reactions, with each reaction defined by its reactants and products, represented as individual chemical species within the graph (Fig. 1). A major advantage of this approach is its capacity to manage "multiple properties" of species and reactions that may originate from various data sources. The unique identification of each species and reaction in the graph

5

permits precise linkage of species and reaction properties to their origin. This is crucial for resolving ambiguities and ensuring data consistency. This structure is also particularly beneficial when designing a new chemical kinetics model because it permits the selection of the most suitable thermodynamic models for species or kinetic models for reactions, based on specific reaction conditions. It allows for accurate comparisons and selection among various models, ensuring a more accurate and context-sensitive application of these models, which is vital for understanding and predicting chemical reactions accurately.

Moreover, the Semantic Web's framework facilitates the alignment of common elementary reactions across different mechanisms, offering a unified view of kinetic data. This system replaces traditional data formats, like Chemkin files, with a dynamic graph structure, where each data point becomes a node in the graph. This enhancement not only simplifies navigation and updates within the vast array of chemical kinetics information but also supports advanced querying capabilities. These features are essential for discovering new patterns and relationships, thus pushing the analysis of chemical kinetics towards a more integrated and insightful approach.
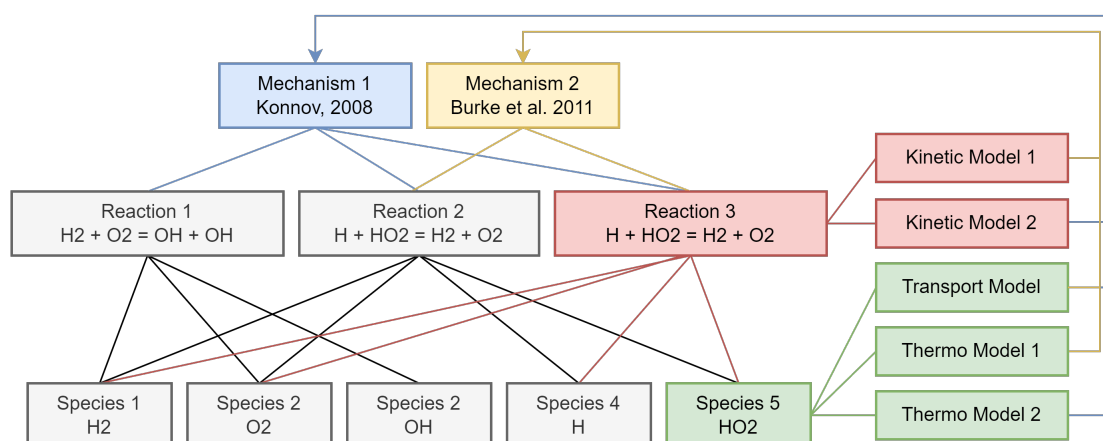
Figure 1: *Illustration of the graph-like nature of the information in chemical kinetics.*

3 The World Avatar infrastructure

The World Avatar (TWA) is a framework that adheres to the FAIR data principles [1]. It is designed to store, process, and analyze chemical data and models. TWA distinguishes itself with a structured, layered architecture, which includes ontologies for data organization, software agents for data manipulation, and connections to empirical real-world data, as depicted in Fig. 2.

At its core, TWA features a **middle layer** composed of a knowledge graph (KG) based on semantic web technologies that provides a dynamic platform for organizing, querying,

and traversing the vast, complex chemical data. The data interaction and connection represented in the graph along with their provenance facilitate data discovery and retrieval of interrelated information and ensure accuracy and traceability of diverse data types, enhancing the understanding of chemical processes. The schema of the KG is defined by ontologies that are split into a Terminology component (TBox) for structural definitions and an Assertion component (ABox) for real-world data instances.

The **top layer** is characterized by the deployment of advanced software agents. These agents are capable of operating both locally and as web services to provide tailored solutions for varied applications, enhancing TWA's functionality for specific user needs.

The **bottom layer** establishes a connection with the real world, integrating empirical data for validation, such as sensor measurements in a laboratory setting, and human-machine interfaces like QA systems for chemistry. This ensures the applicability and reliability of TWA-derived insights and models.

Data and software that form the TWA ecosystem are publicly available [5, 6], allowing for community-driven improvements and adaptations.

The following subsections detail the specific components of TWA utilized in the context of chemical kinetics, discussing modifications that have been implemented for this work.
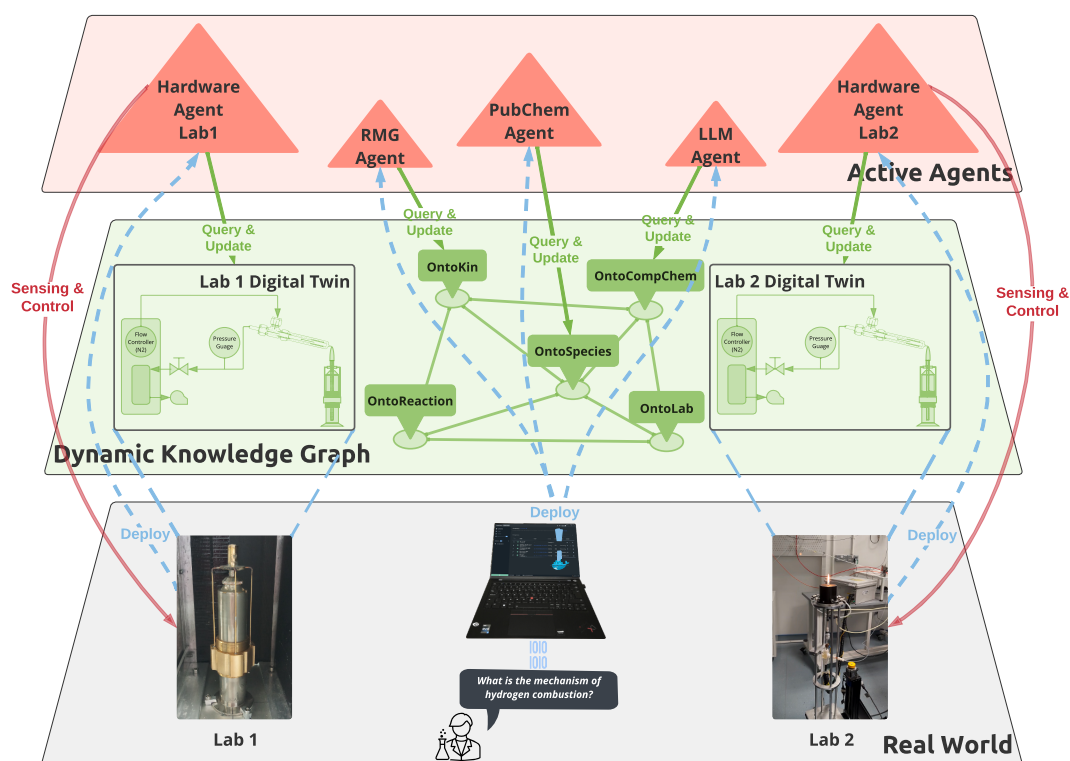
Figure 2: *Illustration of the layered infrastructure of TWA.*

3.1 Knowledge Graphs

In the chemistry domain of TWA KG, several ontologies have been developed to store specific types of information, each addressing a distinct subdomain. This paper concentrates on three subdomains within TWA, each playing a distinct role in enriching the chemical knowledge base as illustrated in Fig. 3.
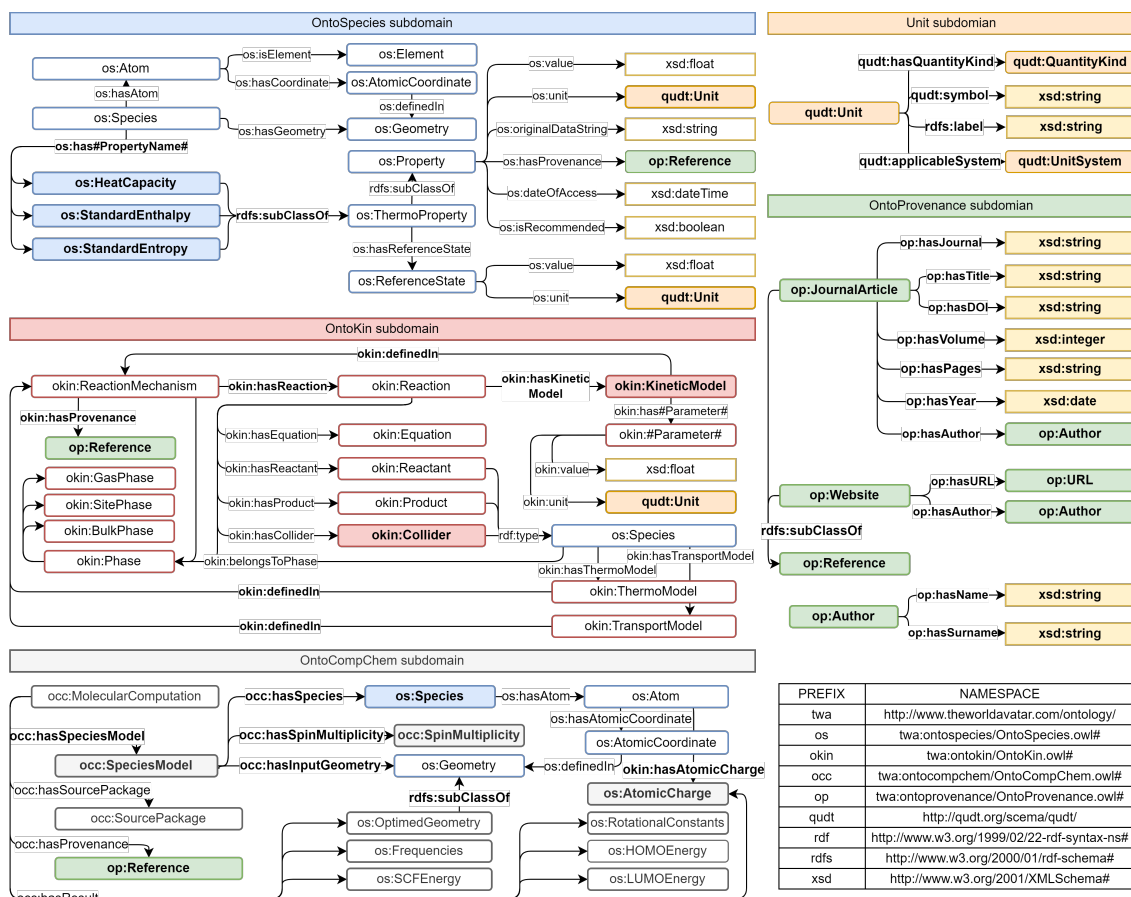

Figure 3: *Simplified schema of TWA TBoxes in the chemistry domain, highlighting interconnectivity and integration of various subdomains shown in different colours. Modifications from previous implementations are distinguished by filled color boxes and bold text.*

OntoSpecies, the core ontology in TWA's chemistry domain (blue boxes in Fig. 3), focuses on species representation and their properties [29]. The majority of data in OntoSpecies is experimental, sourced primarily from PubChem [27]. Despite the uniqueness of certain properties like melting points, OntoSpecies acknowledges the variability that arises when data is collected from diverse sources. As in the PrIMe vision, the data is not organized by source but by merging different sources based on scientific meaning, while meticulously tracking the provenance of each data point [29]. In this work, additional properties (filled blue boxes in Fig. 3) that are crucial for chemical kinetics (`HeatCapacity`, `StandardEnthalpy`, `StandardEntropy`) have been added to the OntoSpecies TBox. These properties are defined as subclasses of `ThermoProperty`

8

and follow the same schema as defined in our previous work [29].

OntoKin is an ontology that represents information about reaction mechanisms, elementary reactions, and kinetic models as presented in the literature, with each entry enriched with provenance details (red boxes in Fig. 3). Compared to its previous implementation [8], it uniquely identifies reactions across different mechanisms based on reactants, products, and possible third bodies. Although a reaction is uniquely identified, it may be associated with different kinetic models used in different reaction mechanisms. This approach extends to species within OntoKin, where different thermochemical and transport models may be applied across various reaction mechanisms. To take this into account, few modifications have been made to OntoKin TBox (filled red boxes in Fig. 3). The `ReactionMechanism` is directly connected to the `ChemicalReaction` by the predicate `hasReaction`. The `ChemicalReaction` has a `KineticModel` that is connected to the `ReactionMechanism` that uses it by the predicate `definedIn`. These connections help to avoid duplicate information and follow the idea presented in Fig. 1, so that the same `KineticModel` can be defined in more than one `ReactionMechanism`. Equivalently, a `Species` that takes part in a `ChemicalReaction` can have more than one `ThermoModel` and `TransportModel` that are linked to the `ReactionMechanism` where they are used in.

OntoCompChem is dedicated to information related to quantum mechanical (QM) calculations [8]. This is particularly crucial for deriving some properties of species, such as heat capacity, especially when experimental data are not available. OntoCompChem bridges the gap between theoretical predictions and empirical data, thereby enhancing the framework's overall accuracy and reliability. The idea of storing calculations and connecting them to the interested species saves time for researchers when the calculation is already available in the ontology. Compared to its previous implementation, species are uniquely identified and connected with their OntoSpecies instance as well as their atoms to avoid any ambiguity.

Lastly, in this work, the addition of unit and provenance subdomains, built upon standard vocabularies and expanded with new units as required, represents an essential step towards standardizing data representation across TWA (orange and green boxes respectively in Fig. 3). Thus, TWA ensures that all data, regardless of its subdomain, adheres to a consistent unit system. This is not only crucial for accurate data representation but also for the interoperability of data between different systems and applications. The standardization of provenance also enhances data interoperability. By embedding this concept directly into the KG, TWA enables the efficient tracking and verification of information sources across its subdomains. This represents a great step forward compared to the widespread use of comment lines in Cantera and Chemkin input files which represented the best practice at the time (see the Appendix).

Together, these ontologies create a synergistic and multifaceted structure within TWA that not only facilitates a comprehensive representation of chemical information but also underscores the depth and complexity inherent in the field of chemical kinetics and computational chemistry. More details on the ontologies' TBoxes can be found at `https://theworldavatar.io/chemistry/documentation`.

3.2 Agents

Software agents within TWA infrastructure fall into two primary categories: those that facilitate the collection, organization, and curation of data, and those that enable the processing and analysis of data.

For the OntoSpecies domain, a specialized agent is tasked with data acquisition from various scientific databases. It retrieves identifiers, experimental properties, and spectral information from PubChem and utilizes ChEBI for chemical classification and usage data [29]. In this work, the population agent has been expanded to extract thermochemical experimental data from the NIST Chemistry WebBook [28]. This comprehensive approach ensures that the OntoSpecies domain is populated with extensive chemical data.

The OntoKin domain employs a population agent that processes data from Chemkin files. The agent firstly translates each chemical species listed in the Chemkin file into an InChI string and links that species with the species IRI as in OntoSpecies, ensuring consistency and integrity of data across domains. The agent then links every reaction in the mechanism to its unique IRI. Finally, with the help of RMG API, it parses the data reported in the Chemkin file and instantiate it in the graph using a SPARQL update. Currently, the data collated by this agent includes four H2/O2 mechanisms taken from literature [4, 19, 22, 24].

The OntoCompChem domain employs a population agent that processes data from Gaussian log files and instantiates the relevant information in the KG.

The second type of agents are used for data augmentation, enhancing the system's capabilities for analysis and prediction. To cite some, if a vibrational analysis of the species exists in the KG, an agent can calculate thermal properties (enthalpy, heat capacity and entropy) from the calculation outputs [8]. We also demonstrated that agents can do sensitivity analysis and calibration, as demonstrated for combustion experiments [2].

In future expansions, we plan to integrate external tools such as RMG for predicting reactive chemistry or others like Cantera, Chemkin, and kinetics, especially as they evolve and new tools emerge [15]. It is important to stress the flexibility of our approach, that structuring data with clear definitions of concepts, units, and provenance, enables easy conversion of knowledge into formats compatible with various tools, that can be then easily integrated in our framework.

3.3 Natural Language Processing

The integration of advanced Natural Language Processing (NLP) techniques for a question-answering (QA) system developed for chemistry within TWA, Marie, marks a significant stride in the realm of chemical kinetics modeling.

The NLP framework in TWA is underpinned by an end-to-end translation approach, leveraging the power of pre-trained text-to-text language models. The system is designed to seamlessly translate questions posed in natural language into SPARQL queries, enabling users to interact with the KG in an intuitive manner. This approach is a departure from traditional methods which relied heavily on hand-crafted templates and were limited in

their scalability and flexibility [31].

In its previous implementation, Marie was limited to the OntoSpecies domain [31]. In this paper, we expanded its functionality to work with additional ontologies like OntoKin and OntoCompChem, achieved by performing multi-task prompted training on both the translation and domain classification tasks. This expansion facilitates a broader range of applications in chemistry-related research and industrial scenarios, making the system an even more versatile tool for data retrieval and analysis.

4 Results and discussion

In the realm of practical application, TWA's advanced querying capabilities, coupled with its comprehensive representation of chemical information, empowers researchers to navigate and resolve the complexities of chemical kinetics modeling. The integration of the QA system, Marie, into TWA infrastructure represents a significant stride towards achieving the vision Frenklach set out in 2007 [9]. As an advanced tool, Marie helps to access TWA capabilities without the need to know the query language SPARQL. In the context of chemical kinetics, Marie is able to answer questions about specific reactions or species as well as full mechanisms and affiliated kinetic or thermodynamic models based on different criteria. In the context of QM calculation, users can request optimised geometries, energies, and other values computed for specific molecules at varying levels of theory and using different basis sets. Marie also handles species-specific queries, like identifying species with particular characteristics or detailing properties of species that belongs to a specific chemical class. Based on its underlying language model, Marie also demonstrates its capabilities in answering more complex questions that require crafting advanced queries across domains and data sets. This can be achieved either by asking for data comparison or chaining simple data retrieval questions, indicating a step closer to having a meaningful conversation with an "AI Scientist".

In this section, we demonstrate the capabilities of Marie in answering questions related to the four H2/O2 reaction mechanism instantiated in the KG as a proof of concept. An illustrative example is presented in Fig. 4, where we compare three different approaches to answering a common research question in chemical kinetics: "Please compare the kinetic models of the reaction H2 + OH = H2O + H across all the mechanisms it appears in." Such queries are crucial for researchers because for common combustion processes, many unique mechanisms and kinetic models are published in the literature and sometimes differ only in a few ways that are hard to compare.

Manual search in publication: The first approach involves a manual search through all publications we want to compare for the specific reaction (H2 + OH = H2O + H). This is a time-consuming process as notations, order of reactants, etc. can differ as shown in the top left of Fig. 4. Moreover, context on the specific kinetic model as well as parameter units need to be analyzed and tracked independently. Lastly, sources need to be tracked closely as well to assess if the same parameter set given in two publications can be interpreted as validations or just using the same underlying model.

Utilizing ChatGPT for Chemkin file parsing: The second approach (shown in the top

Figure 4: *Comparison of data comparison process between manual search of databases, intelligent use of Chemkin-informed ChatGPT, and Marie.*

right of Fig. 4) uses ChatGPT to parse the uploaded Chemkin files related to the publications. Even after a time-consuming additional step of uploading four files and providing context, ChatGPT is unable to identify the reaction in question in any of the mechanisms. Adjusting the prompt to increase precision and rewording the chemical reaction so that it could be found in the file via classic search function does not help either.

Querying Marie: The third approach, asking Marie to compare kinetic models, results in a precise and comprehensive response. Marie efficiently lists all unique kinetic models, including model types, parameter values, and parameter units. It also correctly identify only 3 different kinetic models among the four mechanism instantiated in the KG as two of them use the same one [4, 22]. While the example illustrated in Fig. 4 is limited to Arrhenius models, comparisons across different model types are possible.

Other exemplary use cases can be found in the Appendix.

5   Conclusions and future directions

In this paper, we have presented a proof of concept that showcases the innovative features of TWA in the realm of chemical kinetics modeling and adheres to the PrIMe vision. Central to TWA are its NLP capabilities. By allowing users to interact with the system in natural language, TWA significantly simplifies the process of querying and analyzing chemical data. This feature not only enhances the accessibility and usability of the system but also opens up new avenues for research and practical applications in chemical kinetics.

The vision for TWA and its integrated systems extends beyond the current capabilities

and aligns with perspectives like Bob Kee's on Cantera [17] and Bill Green's on RMG [35], yet seeks to surpass current tool limitations. Key areas for future development include identification of inconsistency, automatic mechanism generation, estimation of rate coefficients, error propagation agents for uncertainty quantification, expanding to other chemistry areas like surface chemistry, integrating with self-driving laboratories and advancing NLP and the reasoning capabilities of Marie. It is crucial to emphasize that the intention is not to reinvent existing tools but to leverage the strengths of established tools, enhancing their utility and effectiveness through seamless integration within the TWA framework, thereby contributing to establishing a global combustion research network as already demonstrated by TWA in other domains [3].

Through these enhancements, TWA and its components aim not only to address the current challenges in the field but also to pave the way for innovative approaches and applications in combustion kinetics and beyond.

Acknowledgements

A Appendix

A.1 Exemplary question-answering use cases with Marie

A.1.1 Data retrieval

One of the primary capabilities of our system is efficient and accurate data retrieval. To demonstrate this, we compare in Fig. 5 three different approaches to answering a common research question in chemical kinetics: "List all the reactions that consume H2O2 in the mechanism by https://doi.org/10.1002/kin.20026." Such queries are crucial for researchers analyzing reaction mechanisms and developing kinetic models, particularly when focusing on specific reactants or products. It is important to note that at this stage, our system identifies a specific mechanism only by its publication DOI.

Figure 5: *Comparison of data retrieval process between manual assessment of publications, intelligent use of Chemkin-informed ChatGPT, and Marie.*

- **Manual search in publication:** The first approach involves a manual search through the publication to find reactions with H2O2 as a reactant. This method is time-consuming and susceptible to human errors, requiring intricate cross-referencing and double-checking.

- **Utilizing ChatGPT for Chemkin file parsing:** The second approach uses ChatGPT to parse a Chemkin file. An additional step of uploading the file and adding some context ("This is a Chemkin input file describing a reaction mechanism") is required, which takes time. While ChatGPT can provide relevant information, its response may miss some or provide too many details.

14

- **Querying Marie:** The third approach, querying Marie with the DOI, results in a precise and comprehensive response. Marie efficiently lists all reactions involving H2O2 as a reactant, eliminating duplicates and omissions. This comparison underscores Marie's advanced capabilities in data retrieval, highlighting its effectiveness and reliability as a tool for detailed chemical kinetics research.

The following prefixes are used for all SPARQL queries provided in this study:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ocape:
↪  <http://www.theworldavatar.com/ontology/ontocape/material/substance/reaction_mechanism.owl#>
PREFIX os: <http://www.theworldavatar.com/ontology/ontospecies/OntoSpecies.owl#>
PREFIX okin: <http://www.theworldavatar.com/ontology/ontokin/OntoKin.owl#>
PREFIX occ: <http://www.theworldavatar.com/ontology/ontocompchem/OntoCompChem.owl#>
```

Mechanism and kinetics:

We can ask for specific reactions or species as well as full mechanisms and affiliated kinetic or thermodynamic models based on different criteria. Four representative examples are given below.

- **Question to Marie:** What are the reactions in which H2 reacts to form OH?

  - **Predicted query:**
    ```
    SELECT ?Reaction WHERE { ?Reaction ocape:hasReactant/skos:altLabel "H2". ?Reaction
    ↪  ocape:hasProduct/skos:altLabel "OH". }
    ```
  - **Post-processed query:**
    ```
    SELECT DISTINCT ?Reaction (SAMPLE(?ReactionEquation) AS ?SampledReactionEquation)
    ↪  WHERE {
        ?Reaction a okin:GasPhaseReaction ; okin:hasEquation ?ReactionEquation .
        ?Reaction ocape:hasReactant/skos:altLabel "H2" .
        ?Reaction ocape:hasProduct/skos:altLabel "OH" .
    }
    GROUP BY ?Reaction
    ```
  - **Result:**

    Show IRIs

Show 10 ⬍ entries	Search:

#	SampledReactionEquation
1	H2 + O <=> H + OH
2	H2 + O2 <=> OH + OH

    Showing 1 to 2 of 2 entries   Previous **1** Next

- **Question to Marie:** What is the kinetic model of the chemical reaction H2O2 + OH = HO2 + H2O involved in the mechanism found in www.osti.gov/servlets/purl/90098-26Ev73/webviewable/?

  - **Predicted query:**
    ```
    SELECT ?KineticModel WHERE { ?Reaction okin:hasEquation "H2O2 + OH = HO2 + H2O".
    ↪  ?Reaction ^okin:hasReaction ?Mechanism. ?Mechanism okin:hasProvenance/op:hasDOI
    ↪  "www.osti.gov/servlets/purl/90098-26Ev73/webviewable/". ?Reaction
    ↪  okin:hasKineticModel ?KineticModel. ?KineticModel okin:definedIn ?Mechanism. }
    ```

– Post-processed query:

```
SELECT DISTINCT ?KineticModel (SAMPLE(?ReactionEquation) AS ?SampledReactionEquation)
↪   ?ModelType ?ActivationEnergyLowValue ?TemperatureExponentHighValue
↪   ?ActivationEnergyUnit ?ArrheniusFactorValue ?TemperatureExponentValue
↪   ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit ?ArrheniusFactorLowUnit
↪   ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
↪   ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
↪   ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
↪   ?T2Value ?ColliderLabel ?ArrheniusFactorUnit WHERE {
    ?Reaction a okin:GasPhaseReaction ; okin:hasEquation ?ReactionEquation .
    ?Reaction okin:hasEquation "H2O2 + OH = HO2 + H2O" .
    ?Reaction ^okin:hasReaction ?Mechanism .
    ?Mechanism okin:hasProvenance/op:hasDOI
    ↪   "www.osti.gov/servlets/purl/90098-26Ev73/webviewable/" .
    ?Reaction okin:hasKineticModel ?KineticModel .
    ?KineticModel a ?KineticModelType .
    BIND (STRAFTER(STR(?KineticModelType), "#") AS ?ModelType)
    OPTIONAL {
        VALUES ?KineticModelType { okin:ArrheniusModel }
        ?KineticModel a okin:ArrheniusModel ; okin:hasActivationEnergy [ okin:value
        ↪   ?ActivationEnergyValue ; okin:unit ?ActivationEnergyUnit ] ;
        ↪   okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorValue ; okin:unit
        ↪   ?ArrheniusFactorUnit ] ; okin:hasTemperatureExponent/okin:value
        ↪   ?TemperatureExponentValue .
    }
    OPTIONAL {
        VALUES ?KineticModelType { okin:MultiArrheniusModel }
        ?KineticModel okin:hasArrheniusModel ?ArrheniusModel .
        ?ArrheniusModel okin:hasActivationEnergy [ okin:value ?ActivationEnergyValue ;
        ↪   okin:unit ?ActivationEnergyUnit ] ; okin:hasArrheniusFactor [ okin:value
        ↪   ?ArrheniusFactorValue ; okin:unit ?ArrheniusFactorUnit ] ;
        ↪   okin:hasTemperatureExponent/okin:value ?TemperatureExponentValue .
    }
    OPTIONAL {
        VALUES ?KineticModelType { okin:ThreeBodyReactionModel okin:LindemannModel
        ↪   okin:TroeModel }
        OPTIONAL {
            ?KineticMode okin:hasCollider [ rdfs:label ?ColliderLabel ;
            ↪   okin:hasEfficiency ?ColliderEfficiency ] .
        }
        ?KineticModel okin:hasArrheniusLowModel ?ArrheniusLowModel .
        ?ArrheniusLowModel okin:hasActivationEnergy [ okin:value
        ↪   ?ActivationEnergyLowValue ; okin:unit ?ActivationEnergyLowUnit ] ;
        ↪   okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorLowValue ; okin:unit
        ↪   ?ArrheniusFactorLowUnit ] ; okin:hasTemperatureExponent/okin:value
        ↪   ?TemperatureExponentLowValue .
        OPTIONAL {
            ?KineticModel okin:hasArrheniusLowModel ?ArrheniusHighModel .
            ?ArrheniusHighModel okin:hasActivationEnergy [ okin:value
            ↪   ?ActivationEnergyHighValue ; okin:unit ?ActivationEnergyHighUnit ] ;
            ↪   okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorHighValue ;
            ↪   okin:unit ?ArrheniusFactorHighUnit ] ;
            ↪   okin:hasTemperatureExponent/okin:value ?TemperatureExponentHighValue .
            }
            OPTIONAL {
                ?KineticModel okin:hasAlpha/okin:value ?AlphaValue ;
                ↪   okin:hasT1/okin:value ?T1Value ; okin:hasT2/okin:value ?T2Value ;
                ↪   okin:hasT3/okin:value ?T3Value .
            }
        }
    }
    ?KineticModel okin:definedIn ?Mechanism .
}
```

```
GROUP BY ?KineticModel ?ModelType ?ActivationEnergyLowValue
↪   ?TemperatureExponentHighValue ?ActivationEnergyUnit ?ArrheniusFactorValue
↪   ?TemperatureExponentValue ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit
↪   ?ArrheniusFactorLowUnit ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
↪   ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
↪   ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
↪   ?T2Value ?ColliderLabel ?ArrheniusFactorUnit
```

- **Result:**

#	SampledReactionEquation	ModelType	ActivationEnergyUnit	ArrheniusFactorValue	TemperatureExponentValue	ActivationEnergyValue	ArrheniusFactorUnit
1	H2O2 + OH <=> H2O + HO2	ArrheniusModel	http://qudt.org/vocab/unit/CAL-PER-MOL	2.4	4.042	-2162.0	cm^3/(mol*s)

Showing 1 to 1 of 1 entries

- **Question to Marie:** For the chemical species that appear in the reaction mechanism found in `https://doi.org/10.1016/j.combustflame.2007.10.024`, what is their thermodynamic model

    – **Predicted query:**

```
SELECT ?ThermoModel WHERE { ?Species okin:belongsToPhase/^okin:hasGasPhase ?Mechanism.
↪   ?Mechanism okin:hasProvenance/op:hasDOI "https:
↪   //doi.org/10.1016/j.combustflame.2007.10.024". ?Species okin:hasThermoModel
↪   ?ThermoModel. ?ThermoModel okin:definedIn ?Mechanism. }
```

    – **Post-processed query:**

```
SELECT DISTINCT ?ThermoModel ?SpeciesLabel ?A1 ?A2 ?A3 ?A4 ?A5 ?A6 ?A7 ?B1 ?B2
↪   ?PolyTminValue ?PolyTminUnit ?PolyTmaxValue ?PolyTmaxUnit ?TminValue ?TminUnit
↪   ?TmaxValue ?TmaxUnit WHERE {
    ?Species a/rdfs:subClassOf* os:Species ; skos:altLabel ?SpeciesLabel .
    ?Species okin:belongsToPhase/^okin:hasGasPhase ?Mechanism .
    ?Mechanism okin:hasProvenance/op:hasDOI "https:
    ↪   //doi.org/10.1016/j.combustflame.2007.10.024" .
    ?Species okin:hasThermoModel ?ThermoModel .
    ?ThermoModel okin:hasPolynomial [ okin:hasA1 ?A1 ; okin:hasA2 ?A2 ; okin:hasA3 ?A3
    ↪   ; okin:hasA4 ?A4 ; okin:hasA5 ?A5 ; okin:hasA6 ?A6 ; okin:hasA7 ?A7 ;
    ↪   okin:hasB1 ?B1 ; okin:hasB2 ?B2 ; okin:hasTmin [ okin:value ?PolyTminValue ;
    ↪   okin:unit ?PolyTminUnit ] ; okin:hasTmax [ okin:value ?PolyTmaxValue ;
    ↪   okin:unit ?PolyTmaxUnit ] ] ; okin:hasTmin [ okin:value ?TminValue ; okin:unit
    ↪   ?TminUnit ] ; okin:hasTmax [ okin:value ?TmaxValue ; okin:unit ?TmaxUnit ] .
    ?ThermoModel okin:definedIn ?Mechanism .
}
```

    – **Result:**

#	SpeciesLabel	A1	A2	A3	A4	A5	A6	A7	B1	B2	PolyTminValue	PolyTminUnit	PolyTmaxValue	PolyTmaxUnit	TminValue	TminUnit	TmaxValue	TmaxUnit
1	H2	3.346003	0.007980425	-1.9477917E-5	2.0159668E-8	-7.3760295E-12	-917.92413	0.6830022	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
2	H2	2.9328306	8.265986E-4	-1.4640057E-7	1.5409805E-11	-6.887661E-16	-813.05583	-1.0243164	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
3	H	2.5	0.0	0.0	0.0	0.0	25473.66	-0.4466084	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
4	H	2.5	0.0	0.0	0.0	0.0	25473.66	-0.4466084	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
5	O2	3.7763456	-0.00269784	9.7403455E-6	-9.5760818E-9	3.2072456E-12	-1063.275	3.6845164	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
6	O2	3.661527	6.5541008E-4	-1.4085422E-7	1.2569096E-11	-1.2169341E-15	-1216.1941	3.4120708	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
7	O	3.1690652	-0.0032151	6.4925558E-6	-5.9875016E-9	2.0687612E-12	28125.846	2.0907684	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
8	O	2.551601	-3.8308546E-5	8.4219746E-10	-4.0128718E-12	-4.1747637E-16	29228.764	4.87617	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
9	OH	3.9919841	-0.0024010667	4.616464E-6	-3.8791635E-9	1.3621655E-12	3375.2727	-0.10399549	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
10	OH	2.8385303	0.0011074129	-2.9400022E-7	4.2069872E-11	-2.4228995E-15	3704.1626	5.8449464	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
11	HO2	4.301788	-0.0047490	2.1157554E-5	-2.4275968E-8	9.2920675E-12	294.80875	3.714701	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
12	HO2	4.172266	0.0018812098	-3.4629398E-7	1.9468015E-11	1.7609153E-16	61.61865	2.9577973	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
13	H2O2	4.3151317	-8.473806E-4	1.7640422E-5	-2.2676295E-8	9.0890025E-12	-17706.744	3.2737331	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
14	H2O2	4.579773	0.00405329	-1.2964473E-6	1.982114E-10	-1.1396876E-14	-18007.178	0.6649707	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
15	H2O	4.198603	-0.0020364018	6.5203417E-6	-5.4879275E-9	1.7719686E-12	-30293.727	-0.84900604	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
16	H2O	2.677038	0.0029731617	-7.7376865E-7	9.4433325E-11	-4.2689995E-15	-29885.893	6.88233	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
17	AR	2.5	0.0	0.0	0.0	0.0	-745.375	4.379875	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
18	AR	2.5	0.0	0.0	0.0	0.0	-745.375	4.379675	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
19	N2	3.5310054	-1.2366098E-4	-5.0299948E-7	2.4353062E-9	-1.4088120E-12	-1046.9763	2.9674704	0.0	0.0	200.0	http://qudt.org/vocab/unit/K	1000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K
20	N2	2.9525764	0.0013969004	-4.9262168E-7	7.8601025E-11	-4.6075525E-15	-923.94847	5.8719877	0.0	0.0	1000.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K	200.0	http://qudt.org/vocab/unit/K	6000.0	http://qudt.org/vocab/unit/K

Showing 1 to 20 of 20 entries

- **Question to Marie:** List all the reactions that consume H2O2 in the mechanism `https://doi.org/10.1002/kin.20026`

17

– **Predicted query:**

```
SELECT ?Reaction WHERE { ?Reaction ^okin:hasReaction ?Mechanism. ?Mechanism
↪   okin:hasProvenance/op:hasDOI "https://doi.org/10.1002/kin.20026". ?Reaction
↪   ocape:hasReactant/skos:altLabel "H2O2". }
```

– **Post-processed query:**

```
SELECT DISTINCT ?Reaction (SAMPLE(?ReactionEquation) AS ?SampledReactionEquation)
↪   WHERE {
    ?Reaction a okin:GasPhaseReaction ; okin:hasEquation ?ReactionEquation .
    ?Reaction ^okin:hasReaction ?Mechanism .
    ?Mechanism okin:hasProvenance/op:hasDOI "https://doi.org/10.1002/kin.20026" .
    ?Reaction ocape:hasReactant/skos:altLabel "H2O2" .
}
GROUP BY ?Reaction
```

– **Result:** See Fig. 5.

Computational chemistry:

We can ask for optimized geometries, energies, and other values computed for specific molecules at varying levels of theory and using different basis sets. Two representative examples are given below.

- **Question to Marie:** What are the HOMO and LUMO energies of O2 at the UB3LYP level of theory?

  – **Predicted query:**

  ```
  SELECT ?HOMOEnergy ?LUMOEnergy WHERE { ?MolecularComputation
  ↪   occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "O2". ?MolecularComputation
  ↪   occ:hasMethodology/occ:hasLevelOfTheory/rdfs:label ?LevelOfTheoryLabel. VALUES
  ↪   ?LevelOfTheoryLabel { "UB3LYP" } ?MolecularComputation occ:hasResult ?HOMOEnergy.
  ↪   ?HOMOEnergy a occ:HOMOEnergy. ?MolecularComputation occ:hasResult ?LUMOEnergy.
  ↪   ?LUMOEnergy a occ:LUMOEnergy. }
  ```

  – **Post-processed query:**

  ```
  SELECT DISTINCT ?BasisSetLabel ?LevelOfTheoryLabel ?HOMOEnergy ?LUMOEnergy
  ↪   ?LUMOEnergyValue ?LUMOEnergyUnit ?HOMOEnergyValue ?HOMOEnergyUnit WHERE {
      ?MolecularComputation occ:hasMethodology/occ:hasBasisSet/rdfs:label ?BasisSetLabel
          ↪   .
      ?LUMOEnergy a occ:LUMOEnergy .
      ?MolecularComputation occ:hasResult ?LUMOEnergy .
      ?LUMOEnergy occ:value ?LUMOEnergyValue ; occ:unit ?LUMOEnergyUnit .
      ?HOMOEnergy a occ:HOMOEnergy .
      ?MolecularComputation occ:hasResult ?HOMOEnergy .
      ?HOMOEnergy occ:value ?HOMOEnergyValue ; occ:unit ?HOMOEnergyUnit .
      VALUES ?LevelOfTheoryLabel { "UB3LYP" }
      ?MolecularComputation occ:hasMethodology/occ:hasLevelOfTheory/rdfs:label
          ↪   ?LevelOfTheoryLabel .
      ?MolecularComputation occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "O2" .
  }
  ```

  – **Result:**

#	BasisSetLabel	LevelOfTheoryLabel	HOMOEnergyValue	HOMOEnergyUnit	LUMOEnergyValue	LUMOEnergyUnit
1	CC-pVTZ	UB3LYP	-0.31285	http://qudt.org/vocab/unit/E_h	0.09788	http://qudt.org/vocab/unit/E_h
2	CC-pVQZ	UB3LYP	-0.31616	http://qudt.org/vocab/unit/E_h	0.08415	http://qudt.org/vocab/unit/E_h
3	CC-pVDZ	UB3LYP	-0.30226	http://qudt.org/vocab/unit/E_h	0.11565	http://qudt.org/vocab/unit/E_h

  Showing 1 to 3 of 3 entries

- **Question to Marie:** Can you provide the molecular structure of H2 calculated using RB3LYP level of theory?

  – **Predicted query:**

  ```
  SELECT ?OptimizedGeometry WHERE { ?MolecularComputation
  ↪   occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "H2". ?MolecularComputation
  ↪   occ:hasMethodology/occ:hasLevelOfTheory/rdfs:label ?LevelOfTheoryLabel. VALUES
  ↪   ?LevelOfTheoryLabel { "RB3LYP" } ?MolecularComputation occ:hasResult
  ↪   ?OptimizedGeometry. ?OptimizedGeometry a occ:OptimizedGeometry. }
  ```

  – **Post-processed query:**

  ```
  SELECT DISTINCT ?BasisSetLabel ?LevelOfTheoryLabel ?OptimizedGeometry ?Atom ?X ?XValue
  ↪   ?XUnit ?Y ?YValue ?YUnit ?Z ?ZValue ?ZUnit WHERE {
      ?MolecularComputation occ:hasMethodology/occ:hasBasisSet/rdfs:label ?BasisSetLabel
      ↪   .
      ?OptimizedGeometry a occ:OptimizedGeometry .
      ?MolecularComputation occ:hasResult ?OptimizedGeometry .
      ?X os:fromGeometry ?OptimizedGeometry ; os:value ?XValue ; os:unit ?XUnit .
      ?Y os:fromGeometry ?OptimizedGeometry ; os:value ?YValue ; os:unit ?YUnit .
      ?Z os:fromGeometry ?OptimizedGeometry ; os:value ?ZValue ; os:unit ?ZUnit .
      ?Atom os:hasXCoordinate ?X ; os:hasYCoordinate ?Y ; os:hasZCoordinate ?Z .
      VALUES ?LevelOfTheoryLabel { "RB3LYP" }
      ?MolecularComputation occ:hasMethodology/occ:hasLevelOfTheory/rdfs:label
      ↪   ?LevelOfTheoryLabel .
      ?MolecularComputation occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "H2" .
  }
  ```

  – **Result:**

#	BasisSetLabel	LevelOfTheoryLabel	XValue	XUnit	YValue	YUnit	ZValue	ZUnit
1	CC-pVQZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	-0.37098	http://qudt.org/vocab /unit/ANGSTROM
2	CC-pVTZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	-0.371456	http://qudt.org/vocab /unit/ANGSTROM
3	CC-pVDZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.380865	http://qudt.org/vocab /unit/ANGSTROM
4	CC-pVQZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.37098	http://qudt.org/vocab /unit/ANGSTROM
5	CC-pVTZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.371456	http://qudt.org/vocab /unit/ANGSTROM
6	CC-pVDZ	RB3LYP	0.0	http://qudt.org/vocab /unit/ANGSTROM	0.0	http://qudt.org/vocab /unit/ANGSTROM	-0.380865	http://qudt.org/vocab /unit/ANGSTROM

Showing 1 to 6 of 6 entries

A.1.2 Complex questions

Based on its underlying language model, Marie also demonstrates its capabilities in answering more complex questions that require crafting advanced queries across domains and data sets. This can be achieved either by asking for data comparison or chaining simple data retrieval questions.

Comparison questions:

- **Question to Marie:** Please compare the kinetic models of the reaction H2 + OH = H2O + H across all the mechanisms it appears in.

  – **Predicted query:**

  ```
  SELECT ?KineticModel ?Mechanism WHERE { ?Reaction okin:hasEquation "H2 + OH = H2O +
  ↪   H". ?Reaction okin:hasKineticModel ?KineticModel. ?KineticModel okin:definedIn
  ↪   ?Mechanism. ?Reaction okin:belongsToPhase/^okin:hasGasPhase ?Mechanism. }
  ```

– **Post-processed query:**

```
SELECT DISTINCT ?KineticModel ?Mechanism (SAMPLE(?ReactionEquation) AS
↪  ?SampledReactionEquation) ?DOI ?ModelType ?ActivationEnergyLowValue
↪  ?TemperatureExponentHighValue ?ActivationEnergyUnit ?ArrheniusFactorValue
↪  ?TemperatureExponentValue ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit
↪  ?ArrheniusFactorLowUnit ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
↪  ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
↪  ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
↪  ?T2Value ?ColliderLabel ?ArrheniusFactorUnit WHERE {
    ?Reaction a okin:GasPhaseReaction ; okin:hasEquation ?ReactionEquation .
    ?Mechanism a okin:ReactionMechanism ; okin:hasProvenance/op:hasDOI ?DOI .
    ?Reaction okin:hasEquation "H2 + OH = H2O + H" .
    ?Reaction okin:hasKineticModel ?KineticModel .
    ?KineticModel a ?KineticModelType .
    BIND (STRAFTER(STR(?KineticModelType), "#") AS ?ModelType)
    OPTIONAL {
    VALUES ?KineticModelType { okin:ArrheniusModel }
    ?KineticModel a okin:ArrheniusModel ; okin:hasActivationEnergy [ okin:value
    ↪  ?ActivationEnergyValue ; okin:unit ?ActivationEnergyUnit ] ;
    ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorValue ; okin:unit
    ↪  ?ArrheniusFactorUnit ] ; okin:hasTemperatureExponent/okin:value
    ↪  ?TemperatureExponentValue .
    }
    OPTIONAL {
        VALUES ?KineticModelType { okin:MultiArrheniusModel }
        ?KineticModel okin:hasArrheniusModel ?ArrheniusModel .
        ?ArrheniusModel okin:hasActivationEnergy [ okin:value ?ActivationEnergyValue ;
        ↪  okin:unit ?ActivationEnergyUnit ] ; okin:hasArrheniusFactor [ okin:value
        ↪  ?ArrheniusFactorValue ; okin:unit ?ArrheniusFactorUnit ] ;
        ↪  okin:hasTemperatureExponent/okin:value ?TemperatureExponentValue . }
    OPTIONAL {
        VALUES ?KineticModelType { okin:ThreeBodyReactionModel okin:LindemannModel
        ↪  okin:TroeModel }
        OPTIONAL {
            ?KineticMode okin:hasCollider [ rdfs:label ?ColliderLabel ;
            ↪  okin:hasEfficiency ?ColliderEfficiency ] .
        }
        ?KineticModel okin:hasArrheniusLowModel ?ArrheniusLowModel .
        ?ArrheniusLowModel okin:hasActivationEnergy [ okin:value
        ↪  ?ActivationEnergyLowValue ; okin:unit ?ActivationEnergyLowUnit ] ;
        ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorLowValue ; okin:unit
        ↪  ?ArrheniusFactorLowUnit ] ; okin:hasTemperatureExponent/okin:value
        ↪  ?TemperatureExponentLowValue .
        OPTIONAL {
            ?KineticModel okin:hasArrheniusLowModel ?ArrheniusHighModel .
            ?ArrheniusHighModel okin:hasActivationEnergy [ okin:value
            ↪  ?ActivationEnergyHighValue ; okin:unit ?ActivationEnergyHighUnit ] ;
            ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorHighValue ;
            ↪  okin:unit ?ArrheniusFactorHighUnit ] ;
            ↪  okin:hasTemperatureExponent/okin:value ?TemperatureExponentHighValue .
            ↪  }
        OPTIONAL {
            ?KineticModel okin:hasAlpha/okin:value ?AlphaValue ; okin:hasT1/okin:value
            ↪  ?T1Value ; okin:hasT2/okin:value ?T2Value ; okin:hasT3/okin:value
            ↪  ?T3Value . }
        }
    }
    ?KineticModel okin:definedIn ?Mechanism .
    ?Reaction okin:belongsToPhase/^okin:hasGasPhase ?Mechanism .
} GROUP BY ?KineticModel ?Mechanism ?DOI ?ModelType ?ActivationEnergyLowValue
↪  ?TemperatureExponentHighValue ?ActivationEnergyUnit ?ArrheniusFactorValue
↪  ?TemperatureExponentValue ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit
↪  ?ArrheniusFactorLowUnit ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
↪  ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
↪  ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
↪  ?T2Value ?ColliderLabel ?ArrheniusFactorUnit
```

– **Result:** See Fig. 4

- **Question to Marie:** Compare all kinetic models of reactions that contain H2O2 in the mechanism https://doi.org/10.1002/kin.20026

    – **Predicted query:**

```
SELECT ?KineticModel WHERE { ?Reaction ocape:hasReactant/skos:altLabel "H2O2".
↪  ?Reaction ^okin:hasReaction ?Mechanism. ?Mechanism okin:hasProvenance/op:hasDOI
↪  "https://doi.org/10.1002/kin.20026". ?Reaction okin:hasKineticModel ?KineticModel.
↪  ?KineticModel okin:definedIn ?Mechanism. }
```

    – **Post-processed query:**

```
SELECT DISTINCT ?KineticModel (SAMPLE(?ReactionEquation) AS ?SampledReactionEquation)
↪  ?ModelType ?ActivationEnergyLowValue ?TemperatureExponentHighValue
↪  ?ActivationEnergyUnit ?ArrheniusFactorValue ?TemperatureExponentValue
↪  ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit ?ArrheniusFactorLowUnit
↪  ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
↪  ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
↪  ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
↪  ?T2Value ?ColliderLabel ?ArrheniusFactorUnit WHERE {
    ?Reaction a okin:GasPhaseReaction ; okin:hasEquation ?ReactionEquation .
    ?Reaction ocape:hasReactant/skos:altLabel "H2O2" .
    ?Reaction ^okin:hasReaction ?Mechanism .
    ?Mechanism okin:hasProvenance/op:hasDOI "https://doi.org/10.1002/kin.20026" .
    ?Reaction okin:hasKineticModel ?KineticModel .
    ?KineticModel a ?KineticModelType .
    BIND (STRAFTER(STR(?KineticModelType), "#") AS ?ModelType)
    OPTIONAL {
        VALUES ?KineticModelType { okin:ArrheniusModel }
        ?KineticModel a okin:ArrheniusModel ; okin:hasActivationEnergy [ okin:value
        ↪  ?ActivationEnergyValue ; okin:unit ?ActivationEnergyUnit ] ;
        ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorValue ; okin:unit
        ↪  ?ArrheniusFactorUnit ] ; okin:hasTemperatureExponent/okin:value
        ↪  ?TemperatureExponentValue .
    }
    OPTIONAL {
        VALUES ?KineticModelType { okin:MultiArrheniusModel }
        ?KineticModel okin:hasArrheniusModel ?ArrheniusModel .
        ?ArrheniusModel okin:hasActivationEnergy [ okin:value ?ActivationEnergyValue ;
        ↪  okin:unit ?ActivationEnergyUnit ] ; okin:hasArrheniusFactor [ okin:value
        ↪  ?ArrheniusFactorValue ; okin:unit ?ArrheniusFactorUnit ] ;
        ↪  okin:hasTemperatureExponent/okin:value ?TemperatureExponentValue .
    }
    OPTIONAL {
        VALUES ?KineticModelType { okin:ThreeBodyReactionModel okin:LindemannModel
        ↪  okin:TroeModel }
        OPTIONAL {
            ?KineticMode okin:hasCollider [ rdfs:label ?ColliderLabel ;
            ↪  okin:hasEfficiency ?ColliderEfficiency ] .
        }
        ?KineticModel okin:hasArrheniusLowModel ?ArrheniusLowModel .
        ?ArrheniusLowModel okin:hasActivationEnergy [ okin:value
        ↪  ?ActivationEnergyLowValue ; okin:unit ?ActivationEnergyLowUnit ] ;
        ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorLowValue ; okin:unit
        ↪  ?ArrheniusFactorLowUnit ] ; okin:hasTemperatureExponent/okin:value
        ↪  ?TemperatureExponentLowValue .
        OPTIONAL {
            ?KineticModel okin:hasArrheniusLowModel ?ArrheniusHighModel .
            ?ArrheniusHighModel okin:hasActivationEnergy [ okin:value
            ↪  ?ActivationEnergyHighValue ; okin:unit ?ActivationEnergyHighUnit ] ;
            ↪  okin:hasArrheniusFactor [ okin:value ?ArrheniusFactorHighValue ;
            ↪  okin:unit ?ArrheniusFactorHighUnit ] ;
            ↪  okin:hasTemperatureExponent/okin:value ?TemperatureExponentHighValue .
            OPTIONAL {
                ?KineticModel okin:hasAlpha/okin:value ?AlphaValue ;
                ↪  okin:hasT1/okin:value ?T1Value ; okin:hasT2/okin:value ?T2Value ;
                ↪  okin:hasT3/okin:value ?T3Value .
            }
        }
```

```
        }
        ?KineticModel okin:definedIn ?Mechanism .
    }
    GROUP BY ?KineticModel ?ModelType ?ActivationEnergyLowValue
    ↪    ?TemperatureExponentHighValue ?ActivationEnergyUnit ?ArrheniusFactorValue
    ↪    ?TemperatureExponentValue ?ArrheniusFactorHighUnit ?ActivationEnergyLowUnit
    ↪    ?ArrheniusFactorLowUnit ?ColliderEfficiency ?ArrheniusModel ?ActivationEnergyValue
    ↪    ?ActivationEnergyHighValue ?ActivationEnergyHighUnit ?AlphaValue ?T3Value ?T1Value
    ↪    ?ArrheniusFactorLowValue ?ArrheniusFactorHighValue ?TemperatureExponentLowValue
    ↪    ?T2Value ?ColliderLabel ?ArrheniusFactorUnit
```

– **Result:**

• **Question to Marie:** Compare zero-point energy when using CC-pVTZ vs CC-pVQZ basis set for Ar.

– **Predicted query:**

```
SELECT ?ZeroPointEnergy WHERE { ?MolecularComputation
↪    occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "Ar". ?MolecularComputation
↪    occ:hasMethodology/occ:hasBasisSet/rdfs:label ?BasisSetLabel. VALUES
↪    ?BasisSetLabel { "CC-pVTZ" "CC-pVQZ" } ?MolecularComputation occ:hasResult
↪    ?ZeroPointEnergy. ?ZeroPointEnergy a occ:ZeroPointEnergy. }
```

– **Post-processed query:**

```
SELECT DISTINCT ?BasisSetLabel ?LevelOfTheoryLabel ?ZeroPointEnergy
↪    ?ZeroPointEnergyValue ?ZeroPointEnergyUnit WHERE {
        ?MolecularComputation occ:hasMethodology/occ:hasLevelOfTheory/rdfs:label
        ↪    ?LevelOfTheoryLabel .
        ?ZeroPointEnergy a occ:ZeroPointEnergy .
        ?MolecularComputation occ:hasResult ?ZeroPointEnergy .
        ?ZeroPointEnergy occ:value ?ZeroPointEnergyValue ; occ:unit ?ZeroPointEnergyUnit .
        VALUES ?BasisSetLabel { "CC-pVTZ" "CC-pVQZ" }
        ?MolecularComputation occ:hasMethodology/occ:hasBasisSet/rdfs:label ?BasisSetLabel
        ↪    .
        ?MolecularComputation occ:hasSpeciesModel/occ:hasSpecies/rdfs:label "Ar" .
    }
```

– **Result:**

Question chains:

Getting a step closer to having a meaningful conversation with an "AI Scientist", Marie can be asked a series of questions in which a query is formulated based on the answer to the previous one. This is particularly relevant as often times, researchers cannot exactly define the question to which they need an answer to progress their studies. To demonstrate this, we compare three different approaches to searching for thermodynamic models and experimental reference data based on an initially unknown mechanism.
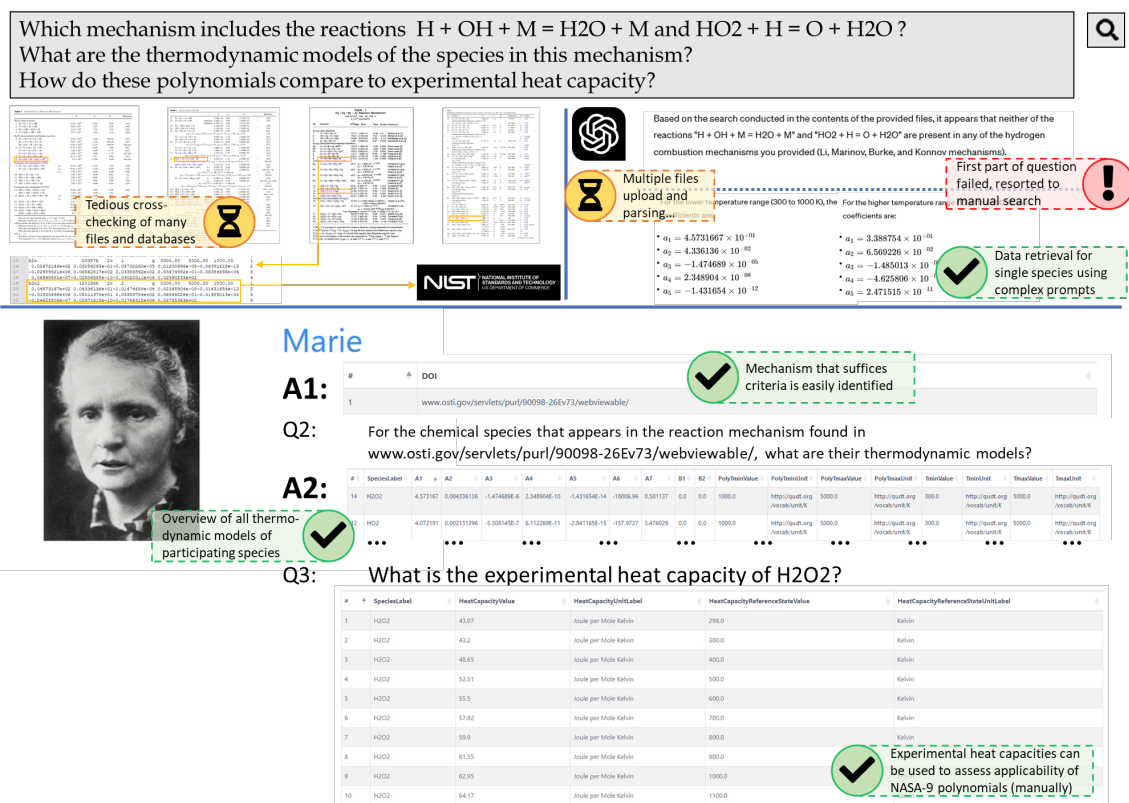
Figure 6: *Comparison of sequential data assessments between manual search of databases, intelligent use of Chemkin-informed ChatGPT, and chained questions to Marie.*

- **Manual search in publication:** The first approach involves a manual search through all publications we want to take into consideration. This is a time-consuming process as notations, orders of reactants, etc. can differ. Based on the findings, thermodynamic models need to be looked up in the supplementary files of the corresponding publication. In order to then get an experimental heat capacity of one of the species involved, an online database such as NIST needs to be consulted.

- **Utilizing ChatGPT for Chemkin file parsing:** The second approach uses ChatGPT to parse the Chemkin files related to the publications. As in the aforementioned use case, ChatGPT is unable to identify reactions in question in any of the mechanisms. It is more successful in the second part of the exercise when fed with

23

a Chemkin input file of thermodynamic models. The last step - obtaining some experimental heat capacities - should be done manually again as ChatGPT proved to be quite inconsistent when asked for such values.

- **Querying Marie:** The third approach, querying Marie in a sequential manner with questions composed around answers obtained in the previous query, results in clearcut answers. This is illustrated in Fig. 6.

The detailed question chain is documented below:

1. **Question to Marie:** Which mechanisms include the reactions H + OH + M = H2O + M and HO2 + H = O + H2O ?

   - **Predicted query:**
     ```
     SELECT ?Mechanism WHERE { ?Mechanism okin:hasReaction/okin:hasEquation "H + OH + M =
     ↪  H2O + M". ?Mechanism okin:hasReaction/okin:hasEquation "HO2 + H = O + H2O". }
     ```
   - **Post-processed query:**
     ```
     SELECT DISTINCT ?Mechanism ?DOI WHERE {
         ?Mechanism a okin:ReactionMechanism ; okin:hasProvenance/op:hasDOI ?DOI .
         ?Mechanism okin:hasReaction/okin:hasEquation "H + OH + M = H2O + M" .
         ?Mechanism okin:hasReaction/okin:hasEquation "HO2 + H = O + H2O" . }
     ```
   - **Result:**

2. **Question to Marie:** For the chemical species that appears in the reaction mechanism found in `www.osti.gov/servlets/purl/90098-26Ev73/webviewable/`, what are their thermodynamic models?

   - **Predicted query:**
     ```
     SELECT ?ThermoModel WHERE { ?Species okin:belongsToPhase/^okin:hasGasPhase ?Mechanism.
     ↪  ?Mechanism okin:hasProvenance/op:hasDOI
     ↪  "www.osti.gov/servlets/purl/90098-26Ev73/webviewable/". ?Species
     ↪  okin:hasThermoModel ?ThermoModel. ?ThermoModel okin:definedIn ?Mechanism. }
     ```
   - **Post-processed query:**
     ```
     SELECT DISTINCT ?ThermoModel ?SpeciesLabel ?A1 ?A2 ?A3 ?A4 ?A5 ?A6 ?A7 ?B1 ?B2
     ↪  ?PolyTminValue ?PolyTminUnit ?PolyTmaxValue ?PolyTmaxUnit ?TminValue ?TminUnit
     ↪  ?TmaxValue ?TmaxUnit WHERE {
         ?Species a/rdfs:subClassOf* os:Species ; skos:altLabel ?SpeciesLabel .
         ?Species okin:belongsToPhase/^okin:hasGasPhase ?Mechanism .
         ?Mechanism okin:hasProvenance/op:hasDOI
         ↪  "www.osti.gov/servlets/purl/90098-26Ev73/webviewable/" .
         ?Species okin:hasThermoModel ?ThermoModel .
         ?ThermoModel okin:hasPolynomial [ okin:hasA1 ?A1 ; okin:hasA2 ?A2 ; okin:hasA3 ?A3
         ↪  ; okin:hasA4 ?A4 ; okin:hasA5 ?A5 ; okin:hasA6 ?A6 ; okin:hasA7 ?A7 ;
         ↪  okin:hasB1 ?B1 ; okin:hasB2 ?B2 ; okin:hasTmin [ okin:value ?PolyTminValue ;
         ↪  okin:unit ?PolyTminUnit ] ; okin:hasTmax [ okin:value ?PolyTmaxValue ;
         ↪  okin:unit ?PolyTmaxUnit ] ] ; okin:hasTmin [ okin:value ?TminValue ; okin:unit
         ↪  ?TminUnit ] ; okin:hasTmax [ okin:value ?TmaxValue ; okin:unit ?TmaxUnit ] .
         ?ThermoModel okin:definedIn ?Mechanism .
     }
     ```

24

- **Result:**

3. **Question to Marie:** What is the experimental heat capacity of H2O2?

- **Predicted query:**

```
SELECT ?HeatCapacity WHERE { VALUES ?Species { "H2O2" } ?Species os:hasHeatCapacity
↪  ?HeatCapacity. }
```

- **Post-processed query:**

```
SELECT DISTINCT ?HeatCapacity ?SpeciesLabel ?Species ?HeatCapacityValue
↪  ?HeatCapacityUnitLabel ?HeatCapacityReferenceStateValue
↪  ?HeatCapacityReferenceStateUnitLabel WHERE {
    ?Species rdf:type os:Species ; rdfs:label ?SpeciesLabel .
    VALUES ?SpeciesIdentifierValue { "H2O2" }
    ?Species ?hasIdentifier [ rdf:type/rdfs:subClassOf os:Identifier ; os:value
    ↪  ?SpeciesIdentifierValue ] .
    ?Species os:hasHeatCapacity ?HeatCapacity .
    ?HeatCapacity os:value ?HeatCapacityValue ; os:unit/rdfs:label
    ↪  ?HeatCapacityUnitLabel .
    OPTIONAL {
        ?HeatCapacity os:hasReferenceState [ os:value ?HeatCapacityReferenceStateValue
        ↪  ; os:unit/rdfs:label ?HeatCapacityReferenceStateUnitLabel ] . }
}
```

- **Result:**

25

A.2 Provenance documentation

Provenance tracking is made much easier within TWA, see Fig. 7

(a) *Provenance as documented in Chemkin files.*

(b) *Provenance as documented in Cantera files.*
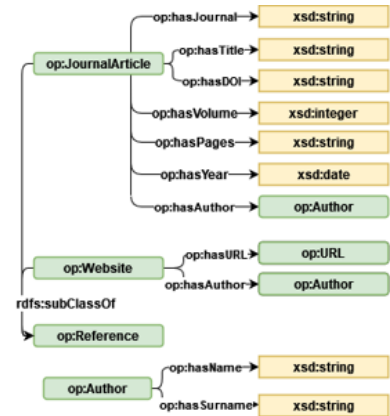
(c) *Provenance tracking via OntoProvenance as part of The World Avatar.*

Figure 7: *Comparison of provenance tracking solutions.*

References

[1] FAIR Principles - GO FAIR. https://www.go-fair.org/fair-principles/, accessed December 2023.

[2] J. Bai, R. Geeson, F. Farazi, S. Mosbach, J. Akroyd, E. J. Bringley, and M. Kraft. Automated calibration of a poly(oxymethylene) dimethyl ether oxidation mechanism using the knowledge graph technology. *J. Chem. Inf. Model.*, 61(4):1701–1717, 2021.

[3] J. Bai, S. Mosbach, C. J. Taylor, D. Karan, K. F. Lee, S. D. Rihm, J. Akroyd, A. A. Lapkin, and M. Kraft. From platform to knowledge graph: Distributed self-driving laboratories, 2023. (preprint).

[4] M. P. Burke, M. Chaos, Y. Ju, F. L. Dryer, and S. J. Klippenstein. Comprehensive H2/O2 kinetic model for high-pressure combustion. *Int. J. Chem. Kinet.*, 44(7): 444–474, 2011.

[5] Cambridge CARES and Computational Modelling Cambridge Ltd. The World Avatar – Chemistry. https://theworldavatar.io/chemistry/documentation, accessed December 2023.

[6] Cambridge CARES and Computational Modelling Cambridge Ltd. The world avatar: A knowledge-graph-based digital twin of the world. https://github.com/cambridge-cares/TheWorldAvatar, accessed December 2023.

[7] C. Cavallotti. Automation of chemical kinetics: Status and challenges. *Proc. Combust. Inst.*, 39:11–28, 2023.

[8] F. Farazi, N. B. Krdzavac, J. Akroyd, S. Mosbach, A. Menon, D. Nurkowski, and M. Kraft. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Comput. Chem. Eng.*, 137:106813, 2020.

[9] M. Frenklach. Transforming data into knowledge-Process Informatics for combustion chemistry. *Proc. Combust. Inst.*, 31 I:125–140, 2007.

[10] C. W. Gao, J. W. Allen, W. H. Green, and R. H. West. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.*, 203:212–225, 2016.

[11] D. G. Goodwin, R. L. Speth, H. K. Moffat, and B. W. Weber. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. https://www.cantera.org, 2021.

[12] W. H. Green. Predictive kinetics: A new approach for the 21st century. *Adv. Chem. Eng.*, 32:1–51, 2007.

[13] W. H. Green. Moving from postdictive to predictive kinetics in reaction engineering. *AIChE J.*, 66:1–16, 2020.

[14] I. Ismail, R. C. Majerus, and S. Habershon. Graph-driven reaction discovery: Progress, challenges, and future opportunities. *J. Phys. Chem. A*, 126:7051–7069, 2022.

[15] M. S. Johnson, H.-W. Pang, A. M. Payne, and W. H. Green. ReactionMechanism-Simulator.jl: A modern approach to chemical kinetic mechanism simulation and analysis. https://doi.org/10.26434/chemrxiv-2023-tj34t, 2023. (preprint).

[16] R. Kee, F. Rupley, J. Miller, M. Coltrin, J. Grcar, E. Meeks, H. Moffat, A. Lutz, G. Dixon-Lewis, M. Smooke, et al. Chemkin collection, release 3.6, reaction design. *Inc., San Diego, CA*, 20(0):0, 2000.

[17] R. J. Kee, S. C. DeCaluwe, and G. S. Jackson. Next-generation chemical-kinetics, transport, and reacting-flow software tools, 2018. (technical report).

[18] K. Kohse-Höinghaus. Combustion, chemistry, and carbon neutrality. *Chem. Rev.*, 123(8):5139–5219, 2023.

[19] A. A. Konnov. Remaining uncertainties in the kinetic mechanism of hydrogen combustion. *Combust. Flame*, 152:507–528, 3 2008. ISSN 0010-2180.

[20] M. Kraft and S. Mosbach. The future of computational modelling in reaction engineering. *Philos. Trans. R. Soc., A*, 368:3633–3644, 2010.

[21] C. E. LaGrotta, Q. Meng, L. Lei, M. C. Barbet, Z. Hong, and M. P. Burke. Resolving discrepancies between state-of-the-art theory and experiment for HO2 + HO2 via multiscale informatics. *J. Phys. Chem. A*, 127:799–816, 1 2023.

[22] J. Li, Z. Zhao, A. Kazakov, and F. L. Dryer. An updated comprehensive kinetic model of hydrogen combustion. *Int. J. Chem. Kinet.*, 36:566–575, 10 2004.

[23] Z. Liu, W. Li, J. Oreluk, A. Hegde, A. Packard, and M. Frenklach. Does a reduced model reproduce the uncertainty of the original full-size model? *Combust. Flame*, 226:98–107, 4 2021.

[24] N. M. Marinov, C. K. Westbrook, and W. J. Pitz. Detailed and global chemical kinetics model for hydrogen (conference) | osti.gov. *Transport Phenomena in Combustion*, 1:1–29, 1995.

[25] J. D. Myers, T. C. Allison, S. Bittner, B. Didier, M. Frenklach, W. H. Green, and et al. A collaborative informatics infrastructure for multi-scale science. *Clust. Comput.*, 8:243–253, 2005.

[26] T. Nagy, Éva Valkó, I. Sedyó, I. G. Zsély, M. J. Pilling, and T. Turányi. Uncertainty of the rate parameters of several important elementary reactions of the h2 and syngas combustion systems. *Combust. Flame*, 162:2059–2076, 5 2015.

[27] National Center for Biotechnology Information. PubChem. https://pubchem.ncbi.nlm.nih.gov/, accessed December 2023.

[28] National Institute of Standards and Technology. NIST Chemistry WebBook. `https://webbook.nist.gov/chemistry/`, accessed December 2023.

[29] L. Pascazio, S. Rihm, A. Naseri, S. Mosbach, J. Akroyd, and M. Kraft. Chemical species ontology for data integration and knowledge discovery. *J. Chem. Inf. Model.*, 63(21):6569–6586, 2023.

[30] O. N. Temkin and D. G. Bonchev. Application of graph theory to chemical kinetics: Part 1. Kinetics of complex reactions. *J. Chem. Educ.*, 69:544–550, 1992.

[31] D. Tran, L. Pascazio, J. Akroyd, S. Mosbach, and M. Kraft. Leveraging text-to-text pre-trained language models for question answering in chemistry. `https://como.ceb.cam.ac.uk/media/preprints/c4e-preprint-313.pdf`, 2023. (preprint).

[32] J. P. Unsleber and M. Reiher. The exploration of chemical reaction networks. *Annu. Rev. Phys. Chem.*, 71:121–142, 2020.

[33] N. M. Vandewiele, K. M. V. Geem, M. F. Reyniers, and G. B. Marin. Genesys: Kinetic model construction using chemo-informatics. *Chemical Engineering Journal*, 207-208:526–538, 10 2012.

[34] T. Varga, T. Turányi, E. Czinki, T. Furtenbacher, and A. Császár. ReSpecTh: A joint reaction kinetics, spectroscopy, and thermochemistry information system. In *Proc. of the 7th Eur. Combust. Meet.*, volume 30, pages 1–5. Budapest, Hungary, 2015.

[35] L. Vereecken, B. Aumont, I. Barnes, J. W. Bozzelli, M. J. Goldman, W. H. Green, and et al. Perspective on mechanism development and structure-activity relationships for gas-phase atmospheric chemistry. *Int. J. Chem. Kinet.*, 50:435–469, 6 2018.