IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

A Novel UX-Based Approach for Ontology Evaluation: Applying Tree Testing to the Agricultural Product Types Ontology

FILIPI MIRANDA SOARES[1,2,3,4], ANTONIO MAURO SARAIVA[2,3], LUÍS FERREIRA PIRES[4], DEBORA PIGNATARI DRUCKER[3,9], KELLY ROSA BRAGHETTO[3,8], LUIZ OLAVO BONINO DA SILVA SANTOS[4,5], DILVAN DE ABREU MOREIRA[3,6], FERNANDO ELIAS CORRÊA[3,7], AND ALEXANDRE CLÁUDIO BOTAZZO DELBEM[3,6]

[1]UMR MISTEA, INRAE, Institut Agro, 2 place Pierre Viala, 34060 Montpellier Cedex 2, France (e-mail: filipi.miranda-soares@inrae.fr)
[2]Polytechnic School, University of São Paulo, Av. Prof. Luciano Gualberto, 158, Butantã, São Paulo, SP 05508-010, Brazil
[3]Center for Artificial Intelligence (C4AI), University of São Paulo, Av. Prof. Luciano Gualberto, 1289, Butantã, São Paulo, SP 05508-020, Brazil
[4]Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Drienerlolaan 5, Enschede, Overijssel 7522 NB, Netherlands
[5]Leiden University Medical Center, Human Genetics, Albinusdreef 2, Leiden, South Holland 2333 ZC, Netherlands
[6]Institute of Mathematics and Computer Science, University of São Paulo, Av. Trabalhador São-carlense, 400, São Carlos, SP 13566-590, Brazil
[7]Luiz de Queiroz College of Agriculture, University of São Paulo, Center for Advanced Studies on Applied Economics, Av. Pádua Dias, 11, Piracicaba, SP 13400-970, Brazil
[8]Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, São Paulo, SP 05508-090, Brazil
[9]Embrapa Digital Agriculture, Av. André Tosello, 209, Campinas, SP 13083-886, Brazil

Corresponding author: Filipi M. Soares (e-mail: filipi.miranda-soares@inrae.fr).

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under grants 2019/07665-4, 2021/15125-0, 2022/08385-8, and 2023/00779-0. This study was also supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001

ABSTRACT This paper introduces a novel approach to evaluating ontologies by integrating user experience (UX) metrics, specifically the Tree Testing protocol, into the validation process. Traditional ontology validation often focuses on verifying competency questions via SPARQL queries, overlooking the critical role of domain specialists in assessing usability and conceptual alignment. To address this gap, we applied tree testing to the Agricultural Product Types Ontology (APTO), tracking specialists' navigation paths as they completed 11 domain-specific tasks. This method provided actionable insights into APTO's usability, revealing structural weaknesses and areas needing refinement. The study involved experts in ontological modeling within the agricultural domain, ensuring feedback was relevant and domain-specific. The findings underscore the value of UX metrics, such as Tree Testing, in identifying user navigation patterns and conceptual misunderstandings. This research illustrates the importance of integrating UX-driven methodologies into ontology design to foster more user-centered and practical knowledge representations for real-world applications. By bridging the gap between technical validation and user-centered evaluation, this work contributes to developing ontologies that are not only technically sound but also intuitive and aligned with the needs of domain experts.

INDEX TERMS Ontology, Usability, User Experience, UX, Validation, Evaluation

I. INTRODUCTION

Ontology testing, as defined in the SABiO methodology by Falbo [1], involves the "verification and validation of the behavior of the operational ontology on a finite set of test cases, against the expected behavior regarding the competency questions" [1, p.8]. Essentially, this means that ontology testing is executed by implementing competency questions as queries in the chosen implementation environment.

However, the SABiO methodology provides a somewhat vague description of validation, particularly regarding the involvement of specialists in the validation process. While an ontology's ability to answer competency questions confirms its syntactical accuracy, it does not necessarily guarantee that the ontology accurately represents the domain or that domain specialists endorse the proposed model. This limitation is not unique to the SABiO methodology. As Casellas [2] notes, most ontology development methodologies offer only a vague description of the involvement of experts in ontology validation. Many ontologies undergo validation informally, often through undocumented discussions with specialists, which risks making the validation process opaque and potentially less reliable.

We argue that more rigorous, transparent, and informative methods are needed for testing and validating ontologies, particularly ontologies of types that function as classification systems. For these ontologies, it is critical that the proposed categorizations (hierarchies) accurately represent the domain and that this precision be validated by specialists. To achieve this, we validated the Agricultural Product Types Ontology (APTO) by using adapted User Experience (UX) metrics, specifically the Tree Testing protocol [3]. This approach offers a comprehensive evaluation of the ontology's usability by allowing the creation of specific tasks and tracking how specialists (users) navigate through an ontology prototype to solve these tasks.

A. THE AGRICULTURAL PRODUCT TYPES ONTOLOGY (APTO)

APTO is a classification ontology designed to standardize the terminology for agricultural product types within the Brazilian commodities market. It addresses the unique challenges posed by Brazil-specific agricultural terms and concepts, many of which lack direct translations into English or carry meanings in Portuguese that are contextually distinct from their English counterparts. This specificity highlights APTO's importance in maintaining the semantic integrity of the Brazilian agricultural terminology while enabling its integration into structured datasets.

APTO follows a modular structure with two key components: the Organism module, linking products to biological origins using the GBIF taxonomy, and the Product Type module, categorizing products by origin (plant, animal, or inorganic) and processing level (raw, processed, or by-product). This dual-faceted design allows APTO to capture complex relationships between products, their sources, and their processing stages, making it adaptable to various use cases, including data integration and semantic web applications.

APTO reuses existing ontologies like Agrovoc[1] and Agrotermos[2] to ensure interoperability and alignment with largely adopted vocabularies. At the same time, APTO addresses gaps in these resources by incorporating over 200 agricultural product types identified in Brazilian agriculture price index datasets, namely Cepea[3], Ipea[4], and Conab[5]. By including terms specific to Brazil's agricultural market, APTO provides a comprehensive framework that enhances data interoperability while maintaining cultural and contextual relevance.

In keeping with the principles of the semantic web, APTO has been implemented in OWL and made available in multiple formats, which can be accessed on AgroPortal[6]. A backup is also maintained on Zenodo [2].

APTO was developed following the SABiO methodology [1]. This methodology encompasses five main phases: (1) Purpose Identification and Requirements Elicitation, (2) Ontology Capture and Formalization, (3) Design, (4) Implementation, and (5) Testing. This paper, however, focuses on the ontology evaluation, which is considered a support process in the SABiO methodology.

II. RELATED WORK

Evaluating ontologies through UX metrics is not a widely adopted practice but has been explored in some studies.

Casellas [2] pioneered the application of usability metrics to ontology evaluation, particularly within the legal domain. This study introduced the idea that usability measures, such as the System Usability Scale (SUS), can effectively assess the practical usability of ontologies from an end-user perspective. Casellas emphasized that while syntactical and structural correctness are important, the actual UX interacting with the ontology is crucial. By considering how domain specialists and users perceive and use the ontology, this method provides a more holistic evaluation, ensuring that the ontology is functionally robust and user-friendly, which is an aspect often overlooked in traditional ontology validation processes.

Other research efforts have concentrated more on the usability of tools and techniques for ontology visualization [4], [5], as well as on the usability of ontology editors and engineering tools [6], [7], rather than directly ontology usability itself. This focus is likely due to the inherent complexity of evaluating ontology usability, particularly when using Task-Oriented Approaches like tree testing or SUS. As noted by Pak [8, p.12], "specifying the characteristics of ontologies is a complicated and time-consuming process; assessing its characteristics is quite subjective." This subjectivity and complexity have made direct evaluations of ontology usability less common despite their importance.

III. USER EXPERIENCE (UX)

The concept of UX emerged in the 1990s as a response to the growing complexity of digital systems and the need for a more holistic approach to understanding how users interact with technology. The term "User Experience" was first popularized by Don Norman, a cognitive scientist and usability expert, during his time at Apple. Norman used the term to emphasize that design should take into account not only usability but also the overall experience of the user,

[1] https://agrovoc.fao.org/browse/agrovoc/en/
[2] https://sistemas.sede.embrapa.br/agrotermos/
[3] https://www.cepea.esalq.usp.br/en
[4] https://www.ipeadata.gov.br/
[5] https://www.conab.gov.br/info-agro/precos
[6] https://w3id.org/APTO#

including emotions, perceptions, and satisfaction [9]. This approach marked a shift from traditional usability — focused on efficiency and effectiveness — towards a broader view of user interaction [9].

A. MEASURING USER EXPERIENCE

Despite its complexity, UX can be assessed through established techniques that provide actionable insights. These insights enable stakeholders to make informed decisions about improvements to products or interfaces. A seminal work in this area by Tullis and Albert [3] offers a robust foundation for understanding different types of UX metrics and the contexts in which they should be applied.

Tullis and Albert [3] begin by differentiating between two primary types of study goals: formative and summative usability.

- **Formative Usability:** This approach is used during the early stages of design to identify potential issues and areas for improvement. It is iterative and diagnostic, aiming to inform the ongoing design process [3].
- **Summative Usability:** In contrast, summative usability is employed after the design is complete to evaluate its overall effectiveness. It is often comparative, used to benchmark against competitors or previous versions of the product [3].

The authors also introduce the concepts of performance and satisfaction as key user goals in UX evaluation.

- **Performance:** This metric focuses on how effectively users can complete tasks using the product. It includes factors such as task success rates, time on task, and error rates [3].
- **Satisfaction:** Satisfaction measures the subjective experience of users, typically assessed through self-reported metrics like surveys and questionnaires. It reflects the user's emotional response to the product [3].

B. TYPES OF USABILITY STUDIES

Tullis and Albert [3] propose ten distinct types of usability studies, each serving a specific purpose:

1) Completing a transaction, which evaluates how easily users can accomplish a specific task, such as purchasing a product online.
2) Comparing products, where multiple products are assessed to determine which offers the best user experience.
3) Evaluating frequent use of the same product, focusing on how users interact with a product they use regularly.
4) Evaluating navigation or information architecture (IA), analyzing how easily users can find information within a system.
5) Increasing awareness measures how effectively a product or feature informs or educates users.
6) Problem discovery, aimed at identifying issues that may hinder the user experience.
7) Maximizing usability for a critical product ensures that essential products or features meet high usability standards.
8) Creating an overall positive user experience, addressing the general perception and emotional response of users.
9) Analyzing the impact of subtle changes, measuring how small modifications affect the user experience.
10) comparing alternative designs, which tests different design approaches to determine which is most effective.

C. EVALUATION METHODS

The authors categorize evaluation methods into three main types:

- **Traditional (Moderated) Usability Tests:** These involve direct observation of users as they interact with the product. Moderators can ask questions and probe deeper into user behaviors.
- **Online (Unmoderated) Usability Tests:** Conducted remotely, these tests allow users to complete tasks in their natural environment without real-time interaction with a moderator.
- **Online Surveys:** These gather self-reported data from users about their experiences and satisfaction levels.

Each of these methods can apply various usability metrics, which Tullis and Albert [3] group into the following categories:

- **Performance Metrics:** These include task success, time on task, error rates, efficiency, and learnability.
- **Issue-Based Metrics:** These focus on identifying specific usability issues encountered by users.
- **Self-Reported Metrics:** These encompass rating scales (such as Likert and Semantic Differential scales), post-task ratings, and post-session ratings, including tools like SUS (as in the case study of Casellas [2], described in the Section II) and the Net Promoter Score (NPS).
- **Behavioral and Physiological Metrics:** These involve tracking user behavior, such as eye tracking (as in the case study of Fu [4], described in the Section II), verbal expressions, and physiological responses like stress or emotional arousal.
- **Special Topics:** These specialized metrics combine different types of metrics, including web analytics (e.g., click-through rates and drop-off rates), A/B testing, card sorting, tree testing, and accessibility data.

IV. METHODS

To evaluate APTO, we selected tree testing as the primary UX measuring protocol. We believe tree testing is particularly suitable for ontology evaluation. It combines various metrics, such as time on task and task success, to assess how users navigate and understand a hierarchy of concepts in an ontology prototype. In UX, a prototype is defined as a preliminary version of a product or system used to explore and evaluate usability. Prototypes can range from low-fidelity to high-fidelity, enabling designers and stakeholders to test ideas and gather feedback before finalizing the design [3].

APTO comprises a total of 581 classes, distributed across two modules: the Product Type module (288 classes) and the Organism module (293 classes). The present study focuses exclusively on the Product Type module, as it contains the conceptual hierarchies most relevant to agricultural commodity classification and user-centered evaluation. The Organism module was not included, since it only includes scientific names.

From the Product Type module, 204 classes were selected for inclusion in the prototype used in the tree testing. The remaining 84 classes were excluded because they directly corresponded to the task target concepts or contained parts of their names. Also, classes that were at the same level as the target concepts were also removed. Including these classes in the prototype would have compromised the validity of the test by revealing the correct answers.

For example, in Task 1, participants were asked to locate the most appropriate upper class for the concept *Whey*. If *Whey* had been included in the prototype, participants could simply navigate directly to it and complete the task without engaging in any meaningful evaluation of the taxonomy.

Usability testing with a prototype allows us to trace user paths through the ontology structure, providing insights into which aspects of the modeling may be unclear or inaccurate from the user's perspective. By analyzing user interactions, we can identify areas for improvement in the ontology design, enhancing its usability and ensuring it aligns more closely with users' domain models.

In the sequel, we discuss the configuration of the tree testing procedure in terms of its goals, participants, and methods.

A. STUDY GOALS

From our perspective as ontology designers, the usability test has been formative. Our primary goal was to detect potential issues and make improvements to a prototype of the ontology before its official release. This objective can be translated into key questions such as: *Can users effectively locate the concepts within the specified ontology hierarchy? What areas require improvement? What insights can we gain from analyzing users' navigation paths within the ontology?* To achieve this goal, we identified "areas of attention" in APTO, which were classified into two types: changes in the original hierarchy of classes imported from *Agrotermos* and *AGROVOC*, and new concepts (not existing in *Agrotermos* or *AGROVOC*) added to the APTO namespace.

From the users' perspective, the usability study was classified as performance-oriented, focusing on evaluating task success rates, time taken to complete tasks, and error rates. Each task involved working with one or more ontology fragments at a time. By "ontology fragments", we refer to branches of the ontology tree, spanning from a broad concept to its most specific sub-concepts. Each task involved navigating one or more fragments to find the superclass to the target concept(s) on that task. These performance metrics offer valuable insights, including:

- **Time on Task:** Longer task completion times may suggest that users perceive the ontology fragments involved in the task as complex or confusing, highlighting areas that could benefit from simplification or reorganization. Conversely, shorter task times may indicate that the ontology fragments are well-structured and easy for users to navigate.
- **Task Success Rates:** High task success rates indicate that the ontology fragments used in a task effectively support users in finding information, validating the logical and intuitive arrangement of categories and relationships. In contrast, low success rates highlight problematic ontology fragments that may require refinement or reorganization.
- **Error Rates:** High error rates suggest that users often misunderstand the categorizations within the ontology fragments used in the task, highlighting a need for restructuring to enhance clarity and usability.

B. PARTICIPANTS

We refer to the participants in this study as "specialist users." While the term is not formally defined in UX literature, it aligns the notion of a "power user" — individuals who utilize advanced features of a system more extensively and effectively than the average user [10]. In the context of our study, specialist users are defined as individuals who possessed two core areas of expertise:

- Experience in ontology engineering.
- Knowledge in the field of Brazilian agriculture, as the ontology is designed for potential users who are specialists working with product types within the Brazilian trading market.

Regarding the number of participants needed for usability testing, Nielsen [11] argued that testing with five users is often sufficient to uncover the majority of usability issues. Norman [9] supports this perspective and emphasizes that a small group of users can reveal the most critical problems, enabling rapid and cost-effective iterations. Tullis and Albert [3] also acknowledge the validity of this approach in many contexts. However, they caution that five users may not always be sufficient, particularly in cases where the user population is highly diverse or where the system being tested is complex and multifaceted. In such scenarios, they recommend increasing the sample size to ensure that the results are representative and that subtler issues, which might only be detected by a broader range of users, are also identified. However, in our case, the diversity of the target population is very limited, since ontology experts working in Brazilian agriculture represent a highly specialized and relatively small community. As such, the recruitment pool is constrained by the low availability of professionals who possess both technical ontology expertise and domain-specific knowledge in agriculture.

Initially, seven participants were recruited for the study. However, the results from one participant were excluded after their task completion time was identified as a statistical outlier (see Section V). The final analysis therefore included data from six valid participants. Given the exploratory nature of the prototype being tested and the specificity of the intended user base, this number is aligned with established UX research guidelines for expert usability studies and was deemed sufficient to yield meaningful insights into the ontology's usability.

C. PRE-TEST

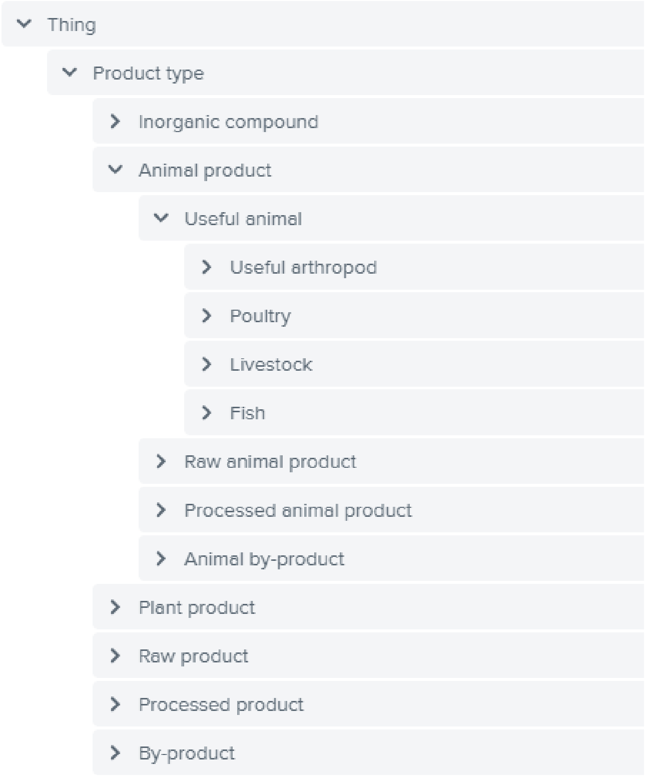
A pre-test was conducted with two specialist users to identify and resolve potential issues in this study protocol. Initially, the study included 12 tasks, but two tasks were found to be very similar and had consistency issues, as noted by one of the pre-test participants. As a result, one task was eliminated from the study. Additionally, the pre-test revealed an oversight where certain correct paths in the prototype were not properly configured, leading to false "wrong answers" being recorded. These issues were corrected prior to the commencement of the study. Pre-test participants did not participate in the actual study.

D. USABILITY STUDY TYPE AND EVALUATION METHOD

Our usability study focused on *Evaluating Navigation and/or Information Architecture*, as participants were tasked with navigating the ontology to complete specific tasks. For the evaluation method, we selected an *Online (Unmoderated) Usability Test*, conducted remotely using the TreeJack tool from Optimal Workshop[7]. This study assessed an interactive prototype of the APTO 'Product Type' module, allowing participants to interact with ontology classes to complete the assigned tasks. The prototype was presented as a hierarchy of concepts, as illustrated in Fig. 1. The full taxonomy of the 'Product Type' module is available in the Data Availability Section (Dataset 1.).

This module is structured as a polyhierarchy — a type of hierarchical structure where a single concept may have multiple parent classes — organized into multiple hierarchical levels based on two classification dimensions: origin (including categories like *Inorganic Compound*, *Animal-Origin Product*, and *Plant-Origin Product*), and type of processing (including categories like *Raw Product*, *Processed Product*, and *By-product*), shown in Fig. 1. This means that tasks could have multiple correct answers within the ontology tree. A brief description was included in the instructions to clarify the meaning of each of these upper classes for the participants.

Eleven tasks were designed, each focused on a specific target concept or group of target concepts. Participants were required to navigate the prototype to identify the most appropriate superclass in the ontology for the target concept(s).



- Thing
  - Product type
    - Inorganic compound
    - Animal product
      - Useful animal
        - Useful arthropod
        - Poultry
        - Livestock
        - Fish
      - Raw animal product
      - Processed animal product
      - Animal by-product
    - Plant product
    - Raw product
    - Processed product
    - By-product

FIGURE 1: APTO Prototype Upper Classes.

The exact names of the target concepts were not visible in the navigation tree, as recommended for this type of protocol [3], requiring participants to explore the categories and make decisions based on the provided descriptions. This setup simulated real-world scenarios where users might need to classify items with limited information. Although the polyhierarchical nature of the ontology allowed for a target concept to belong to more than one superclass, participants were instructed to select only one class as the most appropriate answer for each task. This should reveal navigation preferences within the ontology.

The prototype, instructions, and all supplementary materials used in this study were originally provided to participants in Portuguese and have been translated into English to be included in this paper.

V. RESULTS

Table 1 displays the overall results by participants, including the outlier (participant 3), who was excluded from the analysis. Notably, no task was skipped. Detailed average scores for each task are presented and discussed in Figures 2 to 12[8].

To identify the outlier in task completion times, we used the interquartile range (IQR) measure. Let T be the set of completion times in minutes. The lower quartile Q_1 and upper quartile Q_3 are defined as:

[7] Available at https://www.optimalworkshop.com/product/tree-testing. Accessed: 2024-08-21

[8] Although the tasks are listed sequentially here, participants received them in varying orders, in accordance with the randomization function of TreeJack.

$$Q_1 = 25\text{th percentile of } T, \quad Q_3 = 75\text{th percentile of } T$$

The interquartile range is:

$$\text{IQR} = Q_3 - Q_1$$

An outlier is any time $t \in T$ such that:

$$t > Q_3 + 1.5 \times \text{IQR}$$

Using the observed values:

$$Q_1 = 11.62, \quad Q_3 = 16.19, \quad \text{IQR} = 4.57$$

$$\text{Outlier threshold} = 16.19 + 1.5 \times 4.57 = 23.05 \text{ minutes}$$

Any completion time greater than 23.05 minutes was therefore considered a statistical outlier and excluded from the analysis.

TABLE 1: Participant Task Performance Summary

Part.	Status	Time Taken	Decimal Min.	Question Responses	Tasks Completed	Tasks Skipped	Tasks Successful
1	Compl.	00:18:41	18.68	2	100%	0%	54%
2	Compl.	00:11:35	11.58	2	100%	0%	27%
3	Compl.	06:23:49	383.82	2	100%	0%	45%
4	Compl.	00:08:49	8.82	2	100%	0%	36%
5	Compl.	00:12:44	12.73	2	100%	0%	72%
6	Compl.	00:13:42	13.70	2	100%	0%	90%
7	Compl.	00:11:39	11.65	2	100%	0%	63%

In each chart from Figures 2 to 12, the following scores are included [12]:

- **Success Score:** The percentage of participants who successfully navigated to the correct concept for the task. For each task, there is one or multiple intended correct paths in the ontology.
  - **Direct Success:** Participants who navigated directly to the correct superclass without deviation.
  - **Indirect Success:** Participants who initially navigated down the wrong path, backtracked, and then successfully found the correct superclass.
- **Failure Score:** The percentage of participants who chose an incorrect superclass.
  - **Direct Fail:** Participants who navigated directly to an incorrect superclass without backtracking.
  - **Indirect Fail:** Participants who initially navigated down the wrong path, backtracked, and still ended up at an incorrect superclass.
- **Skip Score:** The percentage of participants who skipped the task before selecting any concept.
  - **Direct Skip:** Participants who clicked the "skip" button without interacting with the tree.
  - **Indirect Skip:** Participants who started navigating through the tree but then chose to skip the task.
- **Directness Score:** The percentage of participants who took a direct path toward their selected answer without backtracking. This score is particularly useful when compared to the success score to determine whether participants were truly successful or if they had to correct their path mid-way.
- **Time Taken:** The average time, in seconds, that participants took to complete the task. This median time is illustrated by the line in the middle of the light blue box on the charts.
- **Overall Score:** A weighted average of the success and directness scores for each task. An overall score of 7 or higher generally indicates good performance.
- **Abandoned:** This score represents instances where a task was neither completed nor skipped, often due to the participant closing the session or timing out. However, no participants abandoned any tasks in this study.

In the following sections, we first present the individual results for each task and subsequently provide a comparative analysis in Section VI. A more detailed analysis of these results is provided in Section VII.

A. TASK 1 - TARGET CONCEPT: *WHEY*

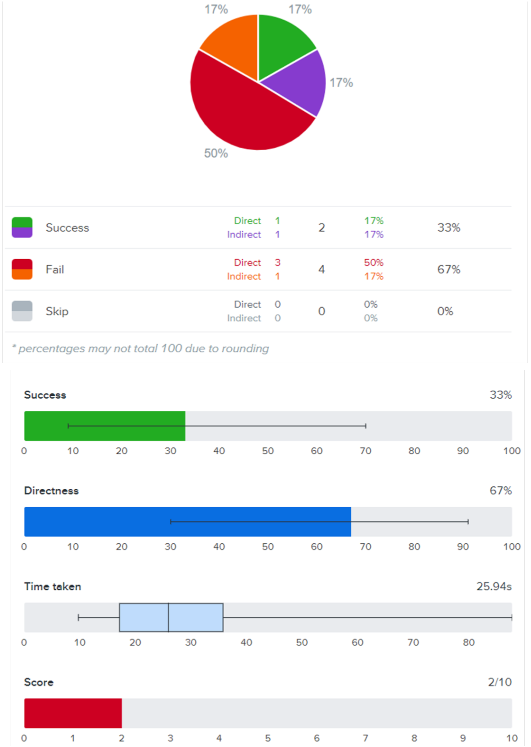
The goal of this task was to determine whether the specialists would categorize *whey* as a by-product, or more specifically, as a *milk by-product*. The interest for this concept validation arises because the classification proposed in APTO differs from that in AGROVOC: while AGROVOC categorizes whey as a processed milk product, APTO classifies it under *milk by-product*. Fig. 2 shows that only two participants selected one of the two correct paths, resulting in a 33% success rate. The directness score of 67% indicates that most participants were confident in their choices, as they did not backtrack. The average time to complete this task was 25.94 seconds. The overall task score was 2, reflecting poor performance on this task.

B. TASK 2 - TARGET CONCEPTS: *SOYBEAN MEAL* AND *WHEAT BRAN*

The objective of Task 2 was to determine whether the participants would categorize *Soybean meal* and *Wheat bran* as *Plant by-products*. This task aimed to validate the modeling of these two new concepts, which were added to the APTO namespace because they were not found in AGROVOC or Agrotermos. Fig. 3 shows that only two participants selected one of the two correct paths, resulting in a 33% success rate. However, the directness score for this task was 33%, indicating that participants frequently backtracked during the task. This backtracking directly influenced the average time to complete the task, which was 65.41 seconds. The overall task score was 2, reflecting poor performance.

C. TASK 3 - TARGET CONCEPTS: *'BOI MAGRO'*, *'BOI GORDO'*, *'VACA GORDA'*, AND *'VACA LEITEIRA'*

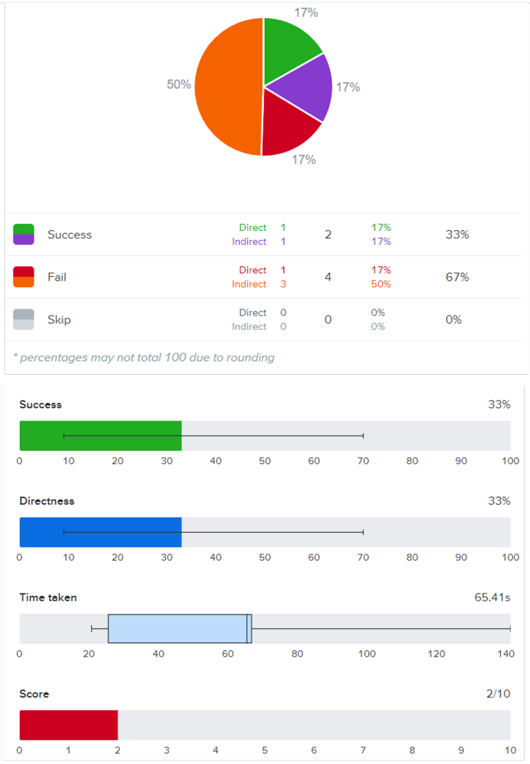
'Boi Magro', *'Boi Gordo'*, *'Vaca Gorda'*, and *'Vaca Leiteira'* are terms used in Brazilian trade to indicate differ-



🟩🟪	Success	Direct Indirect	1 1	2	17% 17%	33%
🟥🟧	Fail	Direct Indirect	3 1	4	50% 17%	67%
⬜	Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

FIGURE 2: **Question:** 'Whey' is the liquid remaining after milk coagulates during cheese production. Rich in proteins, vitamins, and minerals, whey is widely used in the food industry, especially in nutritional supplements, bakery products, and beverages. Which class in the ontology do you consider the most suitable as the upper class for 'whey'?
Correct Paths:
1) Thing > Product type > Animal product > Animal by-product > Milk by-product
2) Thing > Product type > By-product > Animal by-product > Milk by-product



🟩🟪	Success	Direct Indirect	1 1	2	17% 17%	33%
🟥🟧	Fail	Direct Indirect	1 3	4	17% 50%	67%
⬜	Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

FIGURE 3: **Question:** 'Soybean meal' and 'Wheat bran' are obtained from the processing of soybeans and wheat during the production of other primary items, but they are not the intended final product. If you had to choose a single class to group the terms 'Soybean meal' and 'Wheat bran', which class in the ontology would you choose?
Correct Paths:
1) Thing > Product type > Plant product > Plant by-product
2) Thing > Product type > By-product > Plant by-product

ent categories of bovine, such as an ox or cow ready for slaughter (*'Boi Gordo'* and *'Vaca Gorda'*), an ox that still needs to be fattened before slaughter (*'Boi Magro'*), and a cow designated for milk production (*'Vaca Leiteira'*). These concepts are specific to the trading market in Brazil and lack direct translations into English. Additionally, they are not included in Agrotermos or AGROVOC. Therefore, we incorporated these new concepts into the APTO namespace under the superclass *Bovine* (see the complete hierarchy in Fig. 4). This task aimed to validate the modeling of these new concepts within the *Product Types* module. Fig. 4 shows that four participants selected the correct path, resulting in a 67% success rate. The directness score was 50%, indicating that participants moderately backtracked. The average time to complete the task was 35.58 seconds. The overall task score was 4, well below the ideal score of 7.

D. TASK 4 - TARGET CONCEPTS: *FROZEN CHICKEN* AND *CHILLED CHICKEN*

Frozen chicken and *Chilled chicken* refer to chicken meat *in natura*, which undergoes different temperatures of refrigeration to prevent spoilage [13]. Although these meats undergo basic processing (slaughtering and cutting), this processing does not alter the product's composition, so it is still considered a raw product. These two concepts were not found eighter in the Agrotermos nor AGROVOC namespaces; therefore, we added them to the APTO namespace. The aim of this task was to validate the modeling of these new concepts. As shown in Fig. 5, four participants selected one of the four correct paths, resulting in a 67% success rate. The directness score was 67%, indicating that participants had less difficulty navigating the ontology for this task. The average time to complete the task was 35.84 seconds. The overall task score was 5.

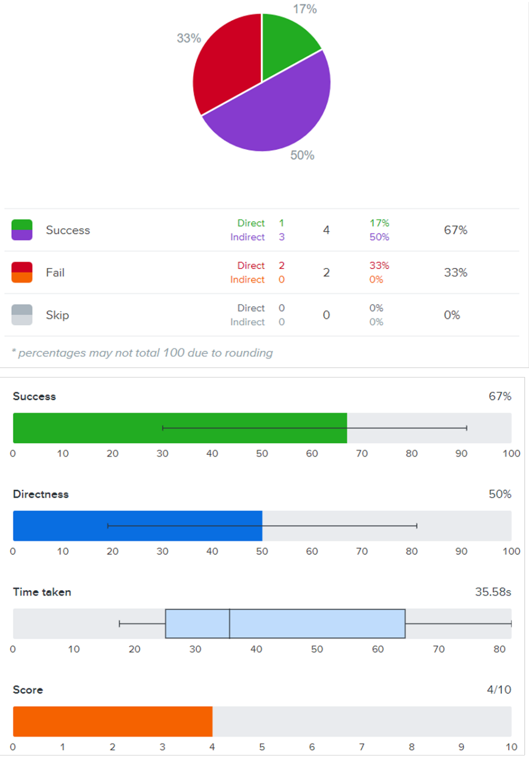


FIGURE 4: **Question:** 'Boi magro' and 'boi gordo' are terms commonly used in the financial market to differentiate between adult cattle ready for slaughter (boi gordo) and those that still need care to reach that point (boi magro). Similarly, 'vaca gorda' refers to a cow ready for slaughter, while 'vaca leiteira' describes a cow intended for milk production. In the ontology, which class do you believe could serve as the 'parent' class capable of grouping all these terms?
Correct Path:
1) Thing > Product type > Animal product > Useful animal > Livestock > Bovine

E. TASK 5 - TARGET CONCEPTS: DAIRY DRINK, YOGURT, POWDERED MILK, PASTEURIZED MILK, UHT MILK, *AND* CHEESE

All the concepts in this task refer to processed milk products. With the exception of *Dairy Drink*, all other concepts were imported from AGROVOC. *Dairy Drink* is the closest translation of the Portuguese term *Bebida láctea*, which has no direct equivalent in English. This concept represents a beverage mix of whey and milk, and is therefore considered a processed product. The aim of this task was to validate the structure incorporated from AGROVOC and the newly added concept. The success rate for this task was significantly higher than in the previous ones, at 83%, as shown in Fig. 6. Additionally, the directness score was 100%, indicating that participants encountered no difficulties navigating to the correct answer. The average time taken was 14.56 seconds, which was also much shorter than in previous tasks. The overall task score was 8, which is considered a good score (above 7) [12]. This suggests participants found this part of the ontology straightforward to navigate.

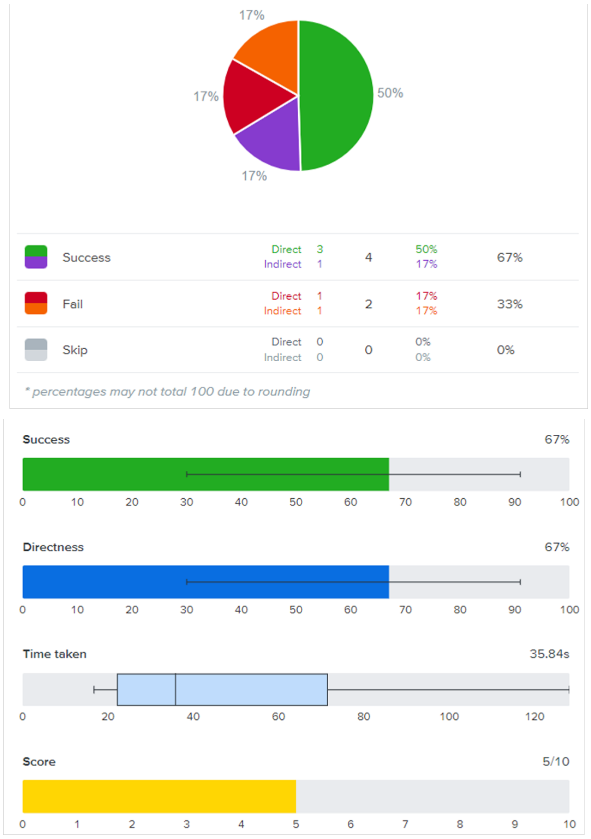
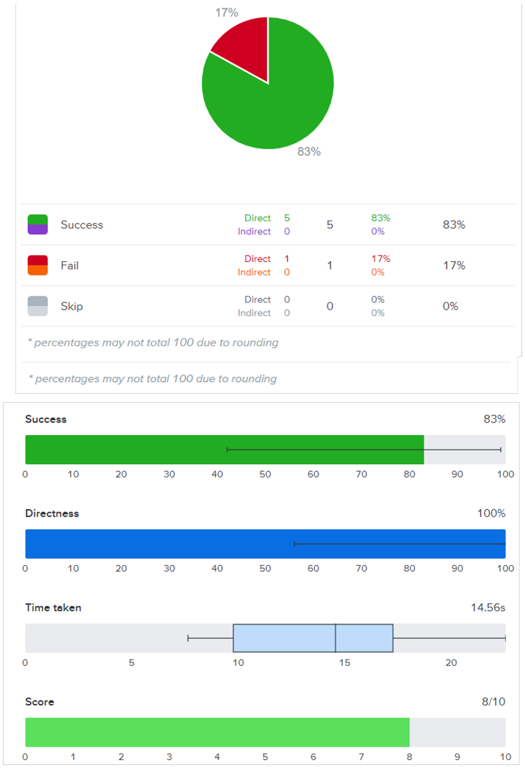


FIGURE 5: **Question:** 'Frozen chicken' and 'chilled chicken' are common terms used to describe slaughtered chicken that undergoes little or no processing beyond slaughter and cleaning and is sent directly from farms to distributors such as cold storage facilities and supermarkets. In the ontology, which class do you believe could serve as the upper class capable of grouping these two terms?
Correct Paths: 1) Thing > Product type > Animal product > Raw animal product > Livestock product > Poultry product > Chicken meat 2) Thing > Product type > Animal product > Raw animal product > Livestock product > Meat > Chicken meat 3) Thing > Product type > Raw product > Raw animal product > Livestock product > Poultry product > Chicken meat 4) Thing > Product type > Raw product > Raw animal product > Livestock product > Meat > Chicken meat

F. TASK 6 - TARGET CONCEPTS: ARAUCARIA, EUCALYPTUS, *AND* PINUS

These concepts representing three types of tree were added to the APTO namespace and are not included in AGROVOC nor Agrotermos in the same way they are represented in APTO. They were categorized as useful plants in APTO, as they are sources of products such as cellulose and wood. The aim of this task was to validate the proposed categorization. The success rate for this task was 17%, as shown in Fig.

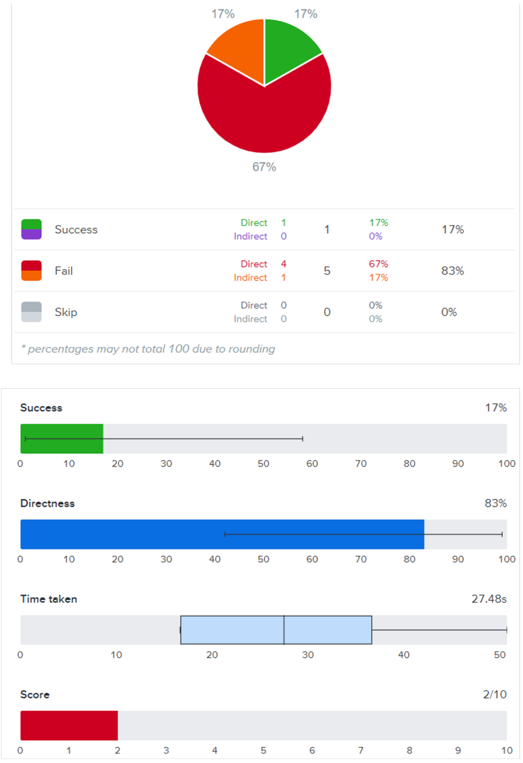


🟩	Success	Direct Indirect	5 0	5	83% 0%	83%
🟥	Fail	Direct Indirect	1 0	1	17% 0%	17%
▨	Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

* percentages may not total 100 due to rounding

FIGURE 6: **Question:** 'Dairy drink', 'yogurt', 'powdered milk', 'pasteurized milk', 'UHT milk', and 'cheese' all have something in common: milk as the main ingredient. Which class in the ontology could be considered the "parent" of these terms?
Correct Paths:
1) Thing > Product type > Animal product > Processed animal product > Milk product
2) Thing > Product type > Processed product > Processed animal product > Milk product



🟩	Success	Direct Indirect	1 0	1	17% 0%	17%
🟥	Fail	Direct Indirect	4 1	5	67% 17%	83%
▨	Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

FIGURE 7: **Question:** 'Araucaria', 'eucalyptus', and 'pinus' are commercial names for trees of the species Araucaria sp., Eucalyptus sp., and Pinus sp., respectively. The cultivation of these species aims at the production of wood and its derivatives, such as cellulose, charcoal, etc. Which upper class in the ontology would best fit the terms araucaria, eucalyptus, and pinus?
Correct Paths:
1) Thing > Product type > Plant product > Useful plant > Cellulose-producing plant
2) Thing > Product type > Plant product > Useful plant > Wood-producing plant

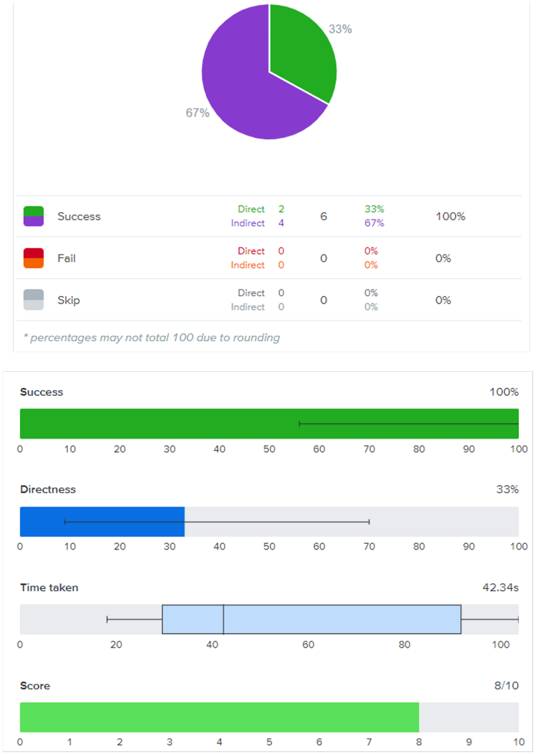
7. This suggests a potentially critical error in the ontology modeling, which prevented users from finding the correct solution. However, the directness score for this task was high, at 83%, and the time on task was low, at 27.48 seconds, indicating that users did not get lost and were confident in the answers they provided. The overall task score was 2, which is considered poor.

G. TASK 7 - TARGET CONCEPTS: *ANDIROBA ALMOND*, *COCOA BEAN*, *CASHEW NUT*, AND *COCONUT*

These products are both edible nuts, as they can be consumed by humans, and oil seeds, as they are also used by the industry to produce oil. The aim of this task was to validate the addition of these new concepts to the APTO schema. Four correct paths were defined for this task, and as shown in Fig. 8, 100% of the participants successfully completed the task. However, the directness score was low, at 33%, and the time on task was high, at 42.34 seconds, indicating that participants found it challenging to locate the correct answer and had to backtrack multiple times. Despite these challenges, the overall score was good, at 8.

H. TASK 8 - TARGET CONCEPTS: *PINEAPPLE PULP*, *AÇAÍ PULP*, *BURITI PULP*, *CERIGUELA PULP*, AND *CUPUAÇU PULP*

These concepts are also new additions to the APTO namespace and did not exist in AGROVOC nor in Agrotermos. When participants navigated the ontology to locate the parent class for these concepts, 67% successfully completed the task. The directness score was moderate, at 50%, indicating that some participants experienced difficulties in finding the correct answer. The time on task was 37.16 seconds, which further supports the hypothesis that participants faced challenges during this task. The overall score was low, at 4, potentially indicating issues in this ontology fragment.



🟩	Success	Direct 2 Indirect 4	6	33% 67%	100%
🟥	Fail	Direct 0 Indirect 0	0	0% 0%	0%
▨	Skip	Direct 0 Indirect 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

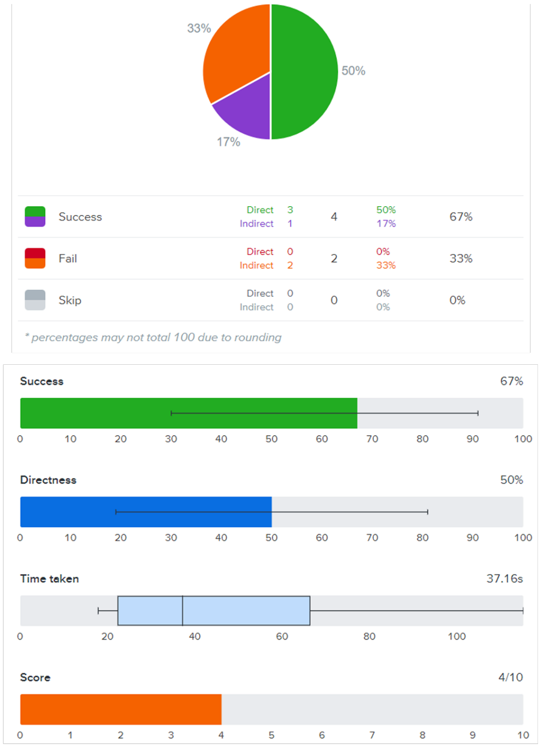
FIGURE 8: **Question:** ‘Andiroba almond’, ‘cocoa bean’, ‘cashew nut’, and ‘coconut’ are examples of oilseeds that can be used for human consumption. In which category of the ontology would you classify them?
Correct Paths:
1) Thing > Product type > Plant product > Raw plant product > Edible nut
2) Thing > Product type > Plant product > Raw plant product > Oil seed
3) Thing > Product type > Raw product > Raw plant product > Edible nut
4) Thing > Product type > Raw product > Raw plant product > Oil seed

I. TASK 9 - TARGET CONCEPTS: *RICE FLAKES* AND *CORN FLAKES*

Rice flakes and *Corn flakes* are processed cereal products obtained from rice and corn, respectively. These concepts are also new additions to the APTO namespace. The success rate for this task was moderate, with only 50% of participants successfully completing it, as shown in Fig. 10. The directness score was also 50%, and the time on task was 29.13 seconds, suggesting that participants may have experienced difficulty navigating the ontology to find the correct paths. Consequently, the overall score was low, at 3.

J. TASK 10 - TARGET CONCEPTS: *CANE SUGAR* AND *ETHANOL*

Cane sugar and *Ethanol* are two processed products derived from sugar cane. Although these concepts are not new and are included in the AGROVOC and Agrotermos namespaces, the superclass defined for them in APTO is new. The term *‘Sucroenergético’*, which has no direct translation to English, is used in Brazil to group these two products derived from sugar cane. The aim of this task was to validate this new class. Only 33% of participants successfully completed this task, as shown in Fig. 11. However, the directness score was high (67%) and the time on task was relatively low (27.74 seconds), indicating that participants were confident in their answers. The resulting overall score was 2, suggesting a potential need for improvements in this ontology fragment.



🟩	Success	Direct 3 Indirect 1	4	50% 17%	67%
🟥	Fail	Direct 0 Indirect 2	2	0% 33%	33%
▨	Skip	Direct 0 Indirect 0	0	0% 0%	0%

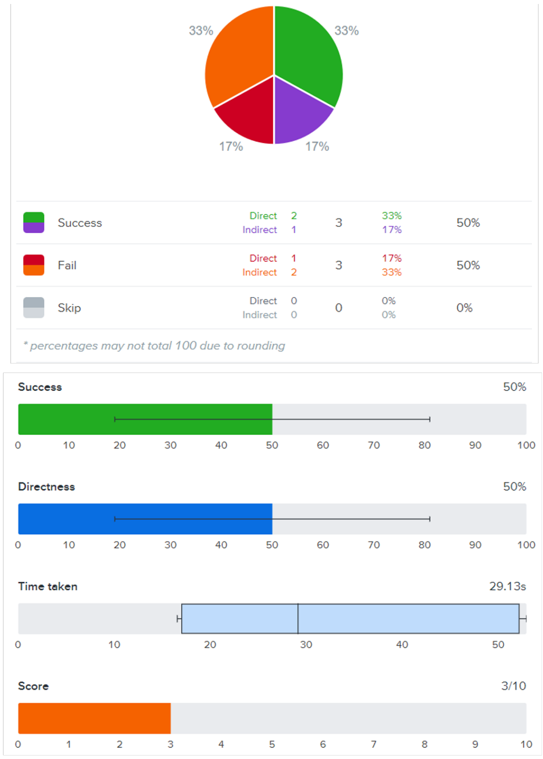
* percentages may not total 100 due to rounding

FIGURE 9: **Question:** ‘Pineapple pulp’, ‘açaí pulp’, ‘buriti pulp’, ‘ceriguela pulp’, and ‘cupuaçu pulp’ are examples of fruit pulps marketed in Brazil. What would be the most suitable upper class in the ontology to group these terms?
Correct Paths:
1) Thing > Product type > Plant product > Processed plant product > Fruit pulp
2) Thing > Product type > Processed product > Processed plant product > Fruit pulp

K. TASK 11 - TARGET CONCEPTS: *BABASSU OIL, MACAUBA OIL, BURITI OIL, COPAIBA OIL, MURUMURU OIL,* AND *PEQUI OIL*

These concepts represent different types of oil seeds that are widely used in oil production in Brazil. In the prototype, there were two correct paths, reflecting the APTO categorization of these classes. These classes do not exist in the Agrotermos or AGROVOC namespaces, so the aim of this task was to validate these new classes added to the APTO namespace. The success rate for this task was considerably high, at 83%,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3595447



Success	Direct Indirect	2 1	3	33% 17%	50%
Fail	Direct Indirect	1 2	3	17% 33%	50%
Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

FIGURE 10: **Question:** ‘Rice flakes’ and ‘corn flakes’ are examples of products processed from cereals. Which class in the ontology could be considered the most suitable parent class for these products?
Correct Paths:
1) Thing > Product type > Plant product > Processed plant product > Cereal product
2) Thing > Product type > Processed product > Processed plant product > Cereal product

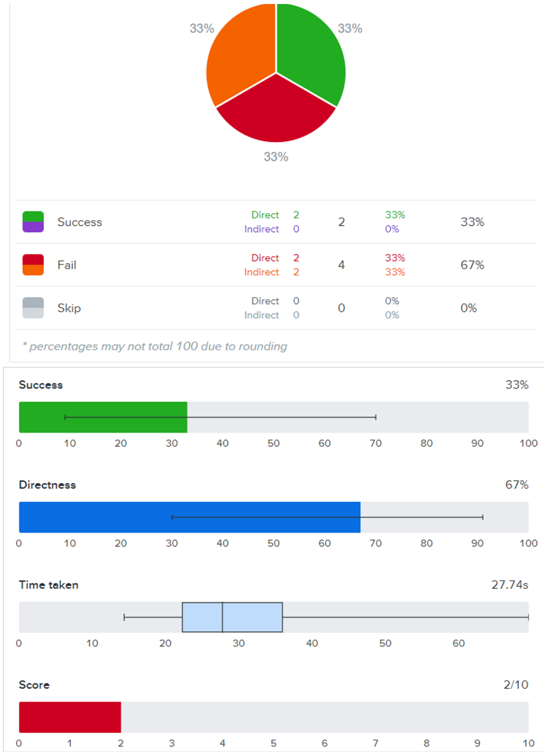
with the same score for directness, as shown in Fig. 12. The average time taken was also low, at 23.54 seconds. The overall score was 7, which is considered good.

VI. COMPARATIVE ANALYSIS OF TASK RESULTS

Based on the results from all 11 tasks (also shown in Table 2), the comparative analysis reveals clear trends, areas of strength, and potential weaknesses in the ontology design, as reflected by participant performance:

A. SUCCESS AND DIRECTNESS

- **High Success and Directness**: Tasks 5 and 11 stand out with success rates of 83%, high directness scores (100% and 83%, respectively), and relatively low completion times. These results indicate that participants found these fragments of the ontology intuitive and easy to navigate, reflecting effective modeling.
- **Moderate Success with Low Directness**: Tasks such as 3, 4, 7, and 8 had success rates of 67% but varied in directness. For example, Task 7 achieved 100% success, but the directness score was only 33%, suggesting that participants struggled with initial navigation and had to backtrack before arriving at the correct solution.
- **Low Success Rates**: Tasks 1, 2, 6, and 10 had success rates of 33% or lower. Task 6 was particularly notable for its poor performance (17% success), despite a high directness score of 83%. This discrepancy suggests that participants confidently chose incorrect answers, high-



Success	Direct Indirect	2 0	2	33% 0%	33%
Fail	Direct Indirect	2 2	4	33% 33%	67%
Skip	Direct Indirect	0 0	0	0% 0%	0%

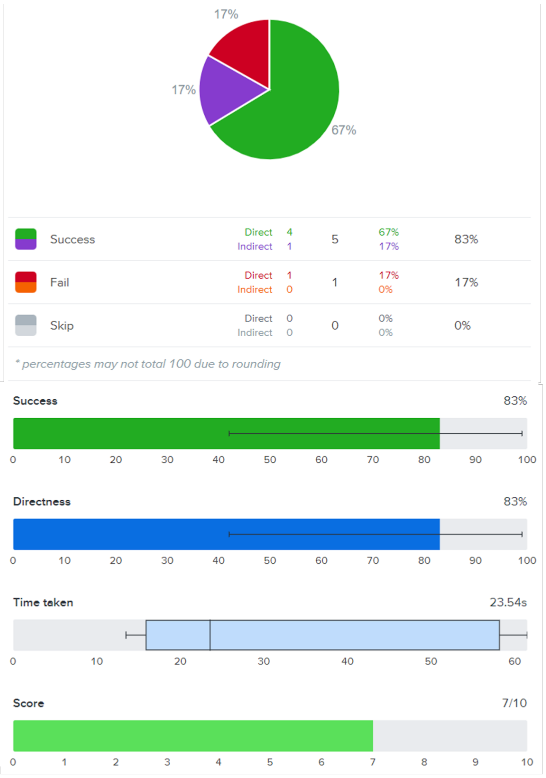
* percentages may not total 100 due to rounding

FIGURE 11: **Question:** ‘Cane sugar’ and ‘ethanol’ are examples of products derived from sugarcane. Which class in the ontology do you consider the most suitable upper class for these products?
Correct Paths:
1) Thing > Product type > Plant product > Processed plant product > Sucroenergético
2) Thing > Product type > Processed product > Processed plant product > Sucroenergético

Task	Success (%)	Directness (%)	Time Taken (s)	Score
Task 1	33	67	25.94	2
Task 2	33	33	65.41	2
Task 3	67	50	35.58	4
Task 4	67	67	35.84	5
Task 5	83	100	14.56	8
Task 6	17	83	27.48	2
Task 7	100	33	42.34	8
Task 8	67	50	37.16	4
Task 9	50	50	29.13	3
Task 10	33	67	27.74	2
Task 11	83	83	23.54	7

TABLE 2: Task Performance Summary

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3595447



🟩 Success	Direct Indirect	4 1	5	67% 17%	83%
🟥 Fail	Direct Indirect	1 0	1	17% 0%	17%
▨ Skip	Direct Indirect	0 0	0	0% 0%	0%

* percentages may not total 100 due to rounding

FIGURE 12: **Question:** 'Babassu oil', 'macaúba oil', 'buriti oil', 'copaiba oil', 'murmuru oil', and 'pequi oil' are examples of vegetable oils extracted from different fruits and seeds native to Brazil. Which class in the ontology do you consider the most suitable "parent" class for these products?
Correct Paths:
1) Thing > Product type > Plant product > Processed plant product > Plant oil
2) Thing > Product type > Processed product > Processed plant product > Plant oil

lighting potential issues in the ontology's structure or categorization logic.

B. TIME ON TASK

- **Quick Completion**: Tasks 5 and 11, with low average times of 14.56 seconds and 23.54 seconds, respectively, indicate clear and straightforward navigation for participants.
- **High Completion Times**: Tasks 2 and 7 had significantly higher times (65.41 and 42.34 seconds, respectively). These longer durations align with their lower directness scores, indicating that participants spent additional time navigating or backtracking.

C. OVERALL PERFORMANCE

- **Best Performance**: Task 5 achieved the highest overall score (8), with perfect directness and high success rates. Similarly, Task 7 also scored 8 due to its perfect success rate, despite the lower directness score.
- **Worst Performance**: Tasks 1, 2, 6, and 10 had overall scores of 2, reflecting significant usability challenges in those fragments of the ontology. Task 6, in particular, indicates a critical need for redesign due to the very low success rate.

D. TRENDS AND INSIGHTS

- **Concept Familiarity**: Tasks involving more familiar or well-established concepts (e.g., processed milk products in Task 5) tended to perform better. In contrast, tasks with less familiar or newly introduced concepts (e.g., tree categorization in Task 6) struggled, suggesting the need for clearer definitions or improved categorization.
- **Ontology Structure and Relationships**: Tasks 7 and 10 highlight the impact of ambiguous or complex relationships between classes. For example, backtracking in Task 7 suggests that participants struggled to distinguish between overlapping categories (e.g., nuts as oil seeds vs. edible products).
- **Task Complexity**: Tasks with multiple correct paths (e.g., Task 11) tended to perform well, as participants had more flexibility in identifying valid answers. However, tasks with stricter paths (e.g., Task 2) presented greater challenges, potentially requiring adjustments to the ontology or the task design.

VII. DETAILED ANALYSIS OF RESULTS

The charts presented in Section V are valuable for understanding the overall performance of participants in each task; however, they provide limited insight into the specific paths participants most frequently chose and the reasons behind their choices. Therefore, in this section, we introduce the so-called Pietrees, which allow us to delve deeper into the details and better comprehend the outcomes for each task. Analyzing these results enabled us to make improvements to APTO and discuss the participants' choices for each task.

A. INTERPRETING PIETREES: INSIGHTS INTO USER NAVIGATION

A Pietree is composed of pies (nodes) representing the concepts in the information architecture (classes in our ontology) and lines depicting the users' navigation paths. The size of a node and the thickness of a line indicate how frequently users navigated through those specific concepts and paths. For example, a thick, green line leading directly from the start to the correct destination suggests that most participants navigated the information architecture with ease, following a clear and intuitive path [14].

Pietrees offer a precise and insightful visualization of participants' navigation paths during a task, allowing to assess how participants build their individual conceptual models for the domain [14].

As suggested by Tullis and Albert [3], positive results in a UX study indicate success but it is by analyzing negative results – where users encounter difficulties – that we can uncover the weaknesses in our information architecture.

Accordingly, this section focuses on discussing the negative results in each task, such as failure and wrong paths, since these highlight the issues in the ontology. The pietree figures presented in this section are also available from Zenodo [1].

B. TASK 1

Fig. 13 showed that one participant incorrectly selected *cow milk* as the superclass for *whey*. However, *cow milk* is classified as a raw product, whereas whey is considered a by-product and for this reason could not be a subclass of *cow milk*.

We believe other relationships between concepts in the ontology could provide clarity, such as `whey ‘derives from’ some ‘Cow milk’`. We believe that if participants had access to these additional relationships, they would have been less likely to misclassify whey under *cow milk*. Another common mistake was classifying whey under *Milk (processed) product*, likely influenced by AGROVOC’s categorization. In AGROVOC, whey is incorrectly modeled as a processed product, whereas it should be classified as a by-product (the class milk by product in AGROVOC has no subclasses).

The rationale for categorizing whey as a by-product in APTO rather than as a processed product is based on its nature as a by-product of cheese, as noted by [15], [16]. In APTO, whey is modeled as a by-product with an added relationship, `‘Whey’ residue_of some ‘Cow cheese’`, to connect whey with its source product. This modeling captures that whey is considered a by-product generated during the production of cheese. The object property `residue_of` is used here to formally express that whey originates from cow cheese. We believe that in a real-world use scenario, the additional relationships provided by APTO would help prevent such misinterpretations of the concept.

C. TASK 2

Fig. 14 reveals that participants frequently backtracked and explored multiple paths during the task. A notable observation is that many participants incorrectly classified *wheat bran* and *soybean meal* under *Cereal product*. This misunderstanding would be somewhat understandable if the task had only involved *wheat bran*, which is derived from wheat flour production. However, the task required participants to identify a common upper class for both *soybean meal* and *wheat bran*, and since soybean is not a cereal, *soybean meal* cannot be classified as a cereal product.

Most participants classified these products as processed products rather than by-products. To address this misconception, we have modeled relationships within the ontology to clarify the connection between processed products and the by-products they generate. Specifically, we used object properties to express that a by-product derives from another product, as follows:

```
Class: ‘Soybean meal’
    SubClassOf: ‘Plant by-product’
    residue_of: exactly 1 ‘Soybean oil’

Class: ‘Wheat bran’
    SubClassOf: ‘Plant by-product’
    residue_of: exactly 1 ‘Wheat flour’
```

In simpler terms, this expression defines that wheat bran is a plant by-product and explicitly models that it results from the processing of wheat flour. The object property residue_of indicates that the by-product is derived from another product.

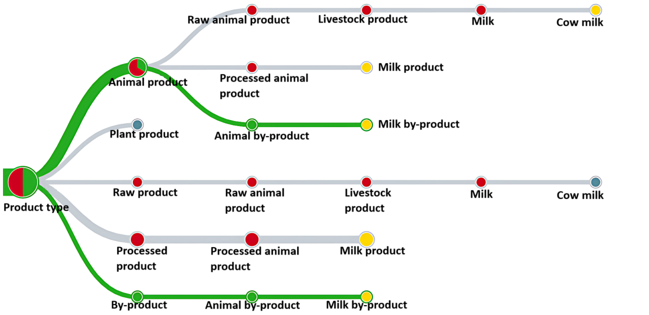


FIGURE 13: Pietree visualization for Task 1 - Classification of *Whey*. This task assessed whether participants could correctly classify whey as a milk by-product. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

D. TASK 3

Participants are rather precise in their navigation choices for this task since Fig. 15 shows that only two paths have been navigated through. Many participants who initially navigated down the wrong path backtracked upon reaching the class *Livestock product*. This suggests that they recognized the task’s requirement to locate the superclass representing the entire animal, rather than a product derived from or part of the animal, which are options under the *Livestock product* class. These participants then corrected their path and successfully identified the correct superclass (*Bovine*).

Some participants classified the concepts under *Beef*, indicating they perceived a relationship between the concepts. Indeed, beef is a product derived from animals of the *Bovine* class. To capture this relationship within the ontology, we modeled it as follows:

```
Class: Beef
    SubClassOf: Meat
    derives_from: some Bovine
```

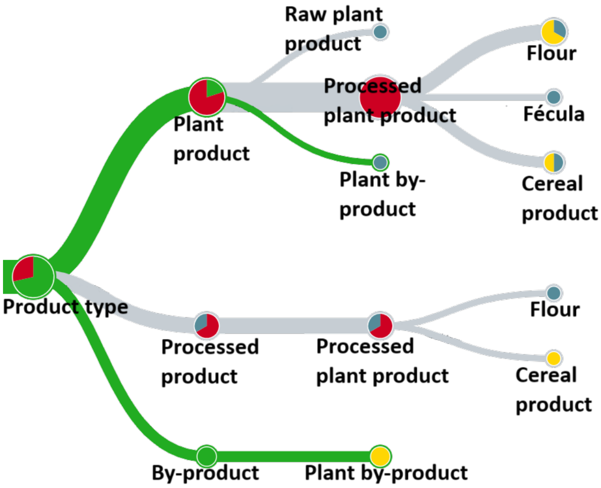


FIGURE 14: Pietree visualization for Task 2 – Classification of *Soybean Meal* and *Wheat Bran*. This task evaluated whether participants would classify these products as plant by-products. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

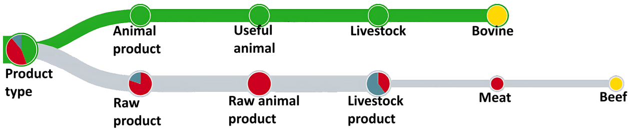


FIGURE 15: Pietree visualization for Task 3 – Classification of *Boi Magro*, *Boi Gordo*, *Vaca Gorda*, and *Vaca Leiteira*. This task assessed whether participants would correctly classify these terms under the superclass *Bovine*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

E. TASK 4

In Task 4, some participants selected *Hive product* as the superclass, while others initially navigated to *Aquaculture product* but backtracked after realizing it was not the correct choice, as Fig. 16 shows. Neither of these classes is related to chicken, as they pertain to products derived from bees and aquatic animals, respectively. We believe this misunderstanding stemmed from the similarity of the upper class names in Portuguese. The correct upper class, *'Produto da avicultura'*, closely resembles the names *'Produto da apicultura'* and *'Produto da aquicultura'*, which have similar syntactical structures and may have caused confusion.

Additionally, some participants navigated to the class *Processed animal product* but backtracked upon realizing it was not the correct upper class. Regarding the correct answers, all four valid paths were used by the participants, indicating that the polyhierarchical modeling of this class aligns well with the participants' conceptual understanding of the domain.

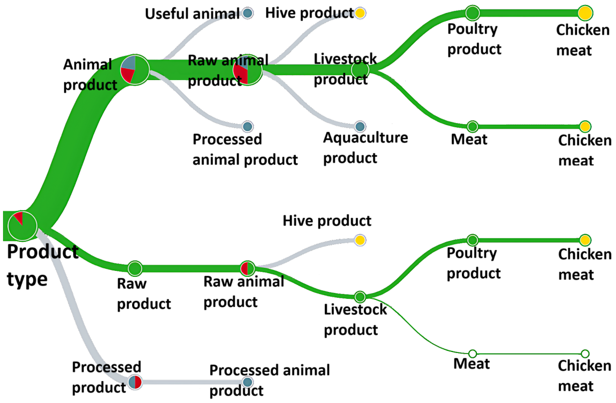


FIGURE 16: Pietree visualization for Task 4 – Classification of *Frozen Chicken* and *Chilled Chicken*. This task evaluated whether participants would correctly classify these items as raw animal products. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

F. TASK 5

Fig. 17 shows that in task 5 participants selected both of the correct paths defined for the task, indicating that the polyhierarchical modeling of this class aligns well with their conceptual understanding of the domain. However, one participant incorrectly selected *Cow milk* as the upper class. It is worth noting that hierarchical relationships in an ontology are often interpreted as "is a" relationships. Thus, concepts like yogurt, dairy drink, and cheese would be incorrectly inferred as "Cheese is a Cow milk", in case the ontology was modeled this way.

The task also included products such as Powdered Milk, Pasteurized Milk, and UHT Milk, which might raise questions about their classification. While these products are indeed types of milk, they undergo extensive processing that significantly alters their composition, classifying them as processed or even ultra-processed products. Since *Cow milk* has *Raw animal product* as its upper class in APTO, these processed milk products do not fit within this category, which represents unprocessed milk. We believe that the associative and hierarchical relationships defined in the full ontology could help prevent such misinterpretations.

In APTO, all the classes in Task 5 are defined as subclasses of *Milk product*. While it might seem natural to add a relationship indicating that a *Milk product* derives from *Cow milk*, this is not always the case. Some milk products are made from the milk of other animals, such as goats, as is the case with certain cheeses. Therefore, the relationship between a specific milk product and the type of milk it comes from is modeled at a more granular level, as shown below:

```
Class: 'Cow cheese'
    SubClassOf: 'Cheese'
    has_ingredient: some 'Cow milk'
```

```
Class: 'Goat cheese'
    SubClassOf: 'Cheese'
    has_ingredient: some 'Goat milk'
```

The relationship is set as 'some' rather than 'exactly 1' because all cheeses contain ingredients in addition to milk.

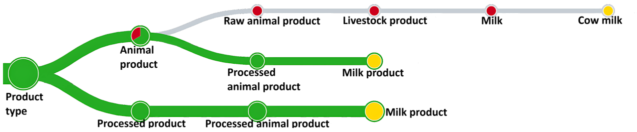


FIGURE 17: Pietree visualization for Task 5 – Classification of *Dairy Drink*, *Yogurt*, *Powdered Milk*, *Pasteurized Milk*, *UHT Milk*, and *Cheese*. This task assessed whether participants would classify these items under the superclass *Milk Product*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

G. TASK 6

Fig. 18 shows a modeling issue in APTO for task 6 regarding the classes *Araucaria*, *Eucalyptus*, and *Pinus*. In APTO, these classes were originally modeled as subclasses of either *Wood-producing plant* or *Cellulose-producing plant*, both of which are subclasses of *Useful plant*. However, most participants classified these classes as subclasses of *Forest product*.

Initially, the *Forest product* class in APTO was designed to group raw products derived from trees, such as wood. We did not consider the trees themselves as forest products. However, after carefully analyzing the participants' choices, we recognized that these trees could indeed be considered forest products, given their role in forestry and their direct economic value as such. This led to revision of APTO to reflect a new hierarchy inferred from the participants' classifications, as below:

```
Class: 'Wood-producing plant'
    SubClassOf: 'Useful plant'
    SubClassOf: 'Forest product'
```

```
Class: 'Cellulose-producing plant'
    SubClassOf: 'Useful plant'
    SubClassOf: 'Forest product'
```

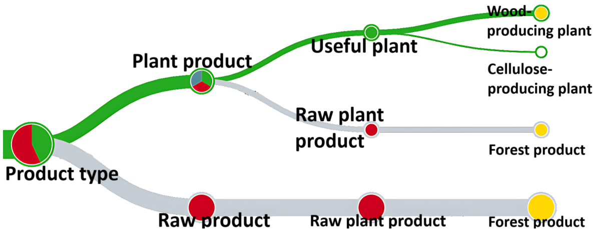


FIGURE 18: Pietree visualization for Task 6 – Classification of *Araucaria*, *Eucalyptus*, and *Pinus*. This task evaluated whether participants would classify these tree species under categories such as *Wood-producing Plant* or *Forest Product*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

Of all the insights obtained from this study, we consider this one to be amongst the most valuable, as it allowed us to identify and correct a fundamental modeling issue in APTO.

H. TASK 7

Task 7 had no incorrect solutions, meaning all users successfully completed the task. However, the directness score of only 33%, as shown in Fig. 8, indicates that while users found the correct path, many initially navigated through incorrect paths before identifying the right one. Fig. 19 provides insight into these wrong paths.

In the upper branch of the Pietree, we observe that some participants initially navigated to *Processed plant product* and *Plant by-product* before backtracking. This is encouraging, as it shows participants correctly recognized that these were not the appropriate superclasses, given that all target concepts in this task represent raw plant products.

In the lower branch of the Pietree, we see that some participants navigated to *Grain* and then backtracked. This reveals a higher-level ontological relationship between the target concepts, which are all types of nuts, and the class *Grain*, as both nuts and grains are types of seeds. Although this relationship is not currently modeled in APTO, it will be considered for inclusion in the next release.

I. TASK 8

Fig. 20 shows a common mistake in Task 8, which was the selection of the superclass *Fruit* as the correct answer.

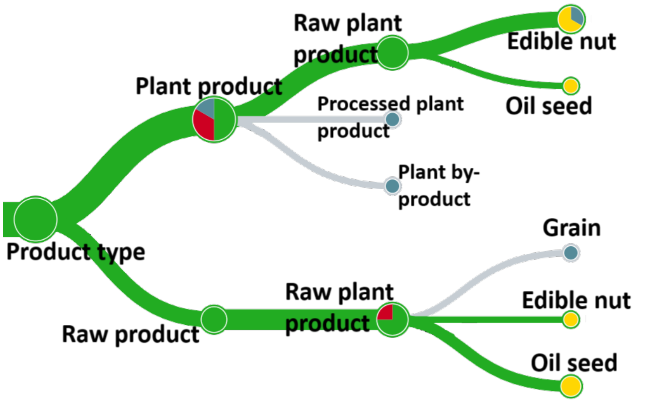


FIGURE 19: Pietree visualization for Task 7 – Classification of *Andiroba Almond*, *Cocoa Bean*, *Cashew Nut*, and *Coconut*. This task assessed whether participants would classify these items as edible nuts or oil seeds. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

However, since *Fruit* is categorized under the superclass *Raw product*, this choice is inconsistent with the nature of *fruit pulp*. *Fruit pulp* is an extract derived from fruit, which means the ontological relationship between fruit and fruit pulp is one of part/whole, rather than a parent/child relationship, making it unsuitable to classify *fruit pulp* directly as subclass of *Fruit*.

Fruit pulps undergo various stages of processing to achieve the desired consistency, flavor, and other properties, which often results in a reduction of beneficial characteristics, such as dietary fiber content [17]. Despite the incorrect selection, it is clear that participants recognized the intrinsic relationship between fruit and fruit pulp. In APTO, we have used object properties to accurately model this relationship between fruit pulp and the fruit from which it is derived, as illustrated for the case of an specific fruit (Assai) below:

```
Class: ‘Assai pulp’
    SubClassOf: ‘Fruit pulp’
    derives_from: exactly 1 Assai
```

This type of relationship represents a recurring modeling problem in APTO and could benefit from the development of an Ontology Design Pattern (ODP) [18]. Specifically, the relationship between derived products (e.g., fruit pulp) and their source (e.g., fruit) requires a consistent and scalable approach that ensures semantic accuracy while minimizing manual effort. An ODP for this purpose could streamline the ontology development process through automation, following these steps:

1) **Template Definition**: Create a reusable template or script that generates subclasses of Fruit pulp for each fruit in APTO.
2) **Automatic Restriction Assignment**: Automate the assignment of the derives_from exactly 1 [Fruit] restriction to each subclass, ensuring consistency in the representation of derivation relationships.
3) **Implementation with Rule-Based Tool**s: Employ a rule-based system or a scripting language (e.g., Python with OWL-RDF libraries or Protégé plugins) to apply the pattern systematically across all relevant fruit types in APTO.

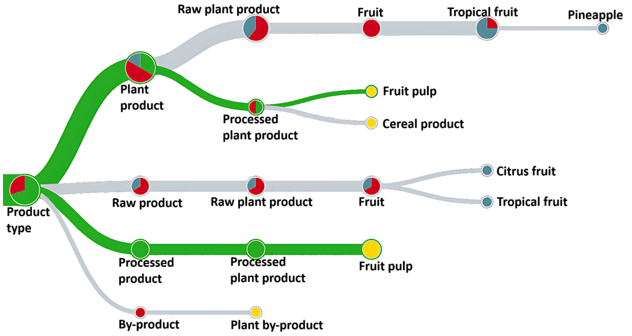


FIGURE 20: Pietree visualization for Task 8 – Classification of *Pineapple Pulp*, *Açaí Pulp*, *Buriti Pulp*, *Ceriguela Pulp*, and *Cupuaçu Pulp*. This task evaluated whether participants would classify these as processed plant products under the class *Fruit Pulp*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

This ODP will be adopted for future updates of APTO to support scalability in modeling this kind of relationship.

J. TASK 9

In Task 9, participants mistakenly selected *Flour*, *Fiber*, and *Plant by-product* as the correct superclasses, as shown in Fig. 21. None of these classes have any relationship with *Corn flakes* and *Rice flakes*. However, none of these classes have a direct relationship with *Corn flakes* or *Rice flakes*. An interesting observation is that most participants followed the correct path up to *Processed plant product*, but then backtracked, unable to find a class they deemed appropriate for the task. To clarify the meaning of these classes and prevent such misunderstandings in the future, we have added definitions to these concepts in APTO.

K. TASK 10

Fig. 22 shows that *Plant by-product* was a popular choice among participants as the correct answer for this task. How-

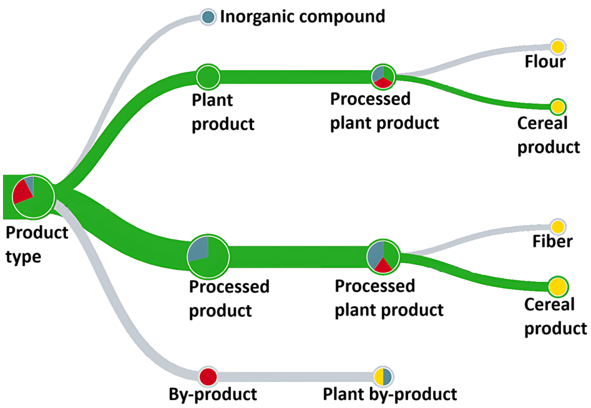


FIGURE 21: Pietree visualization for Task 9 – Classification of *Rice Flakes* and *Corn Flakes*. This task assessed whether participants would identify these items as processed cereal products. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

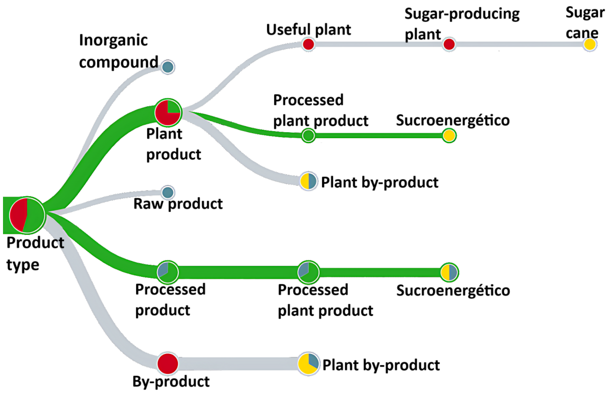


FIGURE 22: Pietree visualization for Task 10 – Classification of *Cane Sugar* and *Ethanol*. This task evaluated whether participants would correctly associate these products with the newly introduced superclass *Sucroenergético*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

ever, neither *Cane sugar* nor *Ethanol* are by-products (they are processed products). Some participants also selected *Sugar cane* as the answer, indicating they recognize a relationship between these products and their source (sugar cane). However, this relationship cannot be classified as a parent/child relationship, since *ethanol* and *cane sugar* are not subtypes of sugar cane, but they are products derived from the processing of sugar cane. To accurately represent this relationship, we modeled it as follows:

```
Class: 'Cane sugar'
    SubClassOf: 'Sugar'
    derives_from: exactly 1 'Sugar cane'

Class: 'Ethanol'
    SubClassOf: 'Alcohol'
    derives_from: some ('Sugar cane'
    or 'Maize')
```

In the case of ethanol, it can also be derived from maize, which is why this was included in the `derives_from` relationship.

L. TASK 11

The only incorrect solution for Task 11 was selecting the class *Oil seed*, as shown in Fig. 23. This choice suggests that the participant understood there is some relationship between these concepts. However, it could not be a parent/child relationship, since oil undergoes several processing stages and cannot be considered a subtype of *Oil seed*, which is categorized as a raw product. In answer to this, in APTO we have modeled the relationships between types of plant oil and the seeds they originate from, as illustrated in the example below:

```
Class: 'Soybean oil'
    SubClassOf: 'Plant oil'
    derives_from: exactly 1 Soybean

Class: 'Murumuru oil'
    SubClassOf: 'Plant oil'
    derives_from: exactly 1 Murumuru
```

Upon analyzing this recurring relationship, we identified it as an Ontology Design Pattern (ODP). Specifically, every subclass of `Plant oil` consistently includes a `derives_from` relationship to specify the source of the oil. For the next version of APTO, we will also automate this ODP.

VIII. DISCUSSION

This study highlights the value of using the Tree Testing protocol to evaluate an ontology module, providing significant insights into participants' navigation paths and choices. These findings contribute to a deeper understanding of how users conceptualize the domain and interact with hierarchical structures in the ontology. However, several limitations emerged during the study, which provide opportunities for improvement and future exploration.

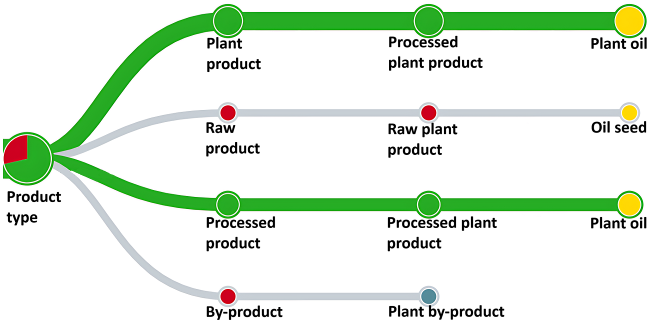


FIGURE 23: Pietree visualization for Task 11 – Classification of *Babassu Oil*, *Macauba Oil*, *Buriti Oil*, *Copaiba Oil*, *Murumuru Oil*, and *Pequi Oil*. This task assessed whether participants would classify these oils under *Plant Oil*. Node colors indicate user actions: green nodes represent navigation through correct paths; red nodes, incorrect paths; blue nodes, backtracking behavior; and the yellow node marks the selected destination. Line colors show navigation direction: green lines represent correct paths; gray lines represent incorrect ones. The size of the nodes and the thickness of the lines reflect the frequency with which participants navigated or selected them.

A. ADVANTAGES AND LIMITATIONS OF THE TREE TESTING PROTOCOL FOR ONTOLOGY EVALUATION

The Tree Testing protocol proved to be an effective tool for identifying usability issues in the ontology, particularly in evaluating hierarchical navigation. The Pietrees, introduced as part of the analysis, offered a detailed visualization of user navigation paths, which helped identify patterns of incorrect answers and backtracking behavior. For example, tasks with lower directness scores highlighted specific areas where participants struggled to find correct paths, revealing weaknesses in the ontology's structure.

However, a key limitation was the inability to fully understand participants' reasoning behind their navigation choices. While navigation paths provided quantitative data, they lacked context about why participants selected certain classes. Although open-ended questions were included in the questionnaire to address this limitation, participants provided minimal feedback. This suggests that complementary methods, such as interviews or think-aloud protocols, could provide richer insights into participants' thought processes and the conceptual challenges they face when navigating the ontology.

Another limitation was the exclusive use of hierarchical relationships in the prototype. Participants did not have access to associative relationships (expressed by object properties), which might have clarified the relationships between concepts. For instance, in Task 1, confusion about whether whey is a milk by-product or a processed product might have been mitigated if participants could see the additional relationship `whey 'derives from' some 'Cow milk'`. Including these relationships in future iterations of the study could improve participants' ability to make accurate classifications and align their understanding with the ontology's design. Thus, a different protocol should be used, since the tree testing protocol only works for hierarchical relationships.

B. TASK-SPECIFIC OBSERVATIONS AND CHALLENGES

Some task designs may have influenced the results. For example, in Task 6, participants struggled to correctly classify *Araucaria*, *Eucalyptus*, and *Pinus*, possibly because the prototype did not reflect participants' mental models of these trees as forest products. Similarly, in Task 5, the classification of *Dairy Drink* was complicated by ongoing debates in Brazil regarding its quality and composition. These examples highlight the need for clearer definitions and targeted testing of new concepts, particularly those that are contentious or culturally specific.

IX. FINAL CONSIDERATIONS

This study sought to evaluate the usability and accuracy of the APTO Product Types module through the application of Tree Testing, a method that allowed us to gain valuable insights into how participants navigate and conceptualize the domain. The findings from this study have not only validated many of the modeling decisions within the ontology but have also highlighted critical areas for improvement. While limitations such as the absence of associative relationships and the lack of qualitative data highlight areas for improvement, this study provides a foundation for refining ontology validation practices.

A. FUTURE DIRECTIONS

Building on these findings, future studies should adopt a mixed-methods approach to ontology evaluation. Combining Tree Testing with qualitative methods — such as interviews or direct observation of users interacting with the ontology — could yield deeper insights into participants' decision-making processes and reasoning patterns.

In parallel, future research should also consider extending usability protocols beyond hierarchical structures to include associative relationships (i.e., object properties). While Tree Testing effectively captures users' understanding of class hierarchies, it does not support the evaluation of other semantic relations, such as `derives_from`, which are essential for interpreting provenance and conceptual derivation. Incorporating visual or interactive mechanisms that make these non-hierarchical relationships explicit may improve classification accuracy and provide a more comprehensive assessment of ontology usability.

Other protocols such as think-aloud and interviews could prove valuable in ontology usability studies, where errors may not arise solely from misunderstandings of hierarchical relationships. These protocols would allow users to verbalize the difficulties they encounter while interacting with the ontology and offer more detailed feedback on their navigation choices. However, these protocols tend to generate noisier data and are generally more complex to analyze than the method we used, as they involve interpreting qualitative responses.

Another avenue for future work is the systematic development and automation of ODPs to address recurring modeling challenges in APTO. For example, automating the assignment of `derives_from` relationships for derived products could streamline ontology development while ensuring semantic accuracy. Expanding the ontology's scope to include such patterns would also facilitate its integration with other vocabularies and enhance its usability for domain experts.

Finally, although this study focused on a specific ontology, APTO, which was designed for the Brazilian agricultural domain, the Tree Testing-based evaluation protocol has the potential to be applied in other domains. We find this protocol especially suited for ontologies that feature rich hierarchical structures. However, its effectiveness may vary depending on the depth, balance, and complexity of the class hierarchy of the ontologies being evaluated. Ontologies with relatively flat structures or heavily reliant on associative (non-hierarchical) relationships may require adaptations or complementary evaluation methods. Future research could explore the applicability of this approach across diverse ontology types and domains, helping establish broader methodological guidelines.

DATA AVAILABILITY

1) Soares, F. M. (2024). Usability Evaluation of the Agriculture Product Types Ontology (APTO) [Dataset]. Version 2. Zenodo. https://doi.org/10.5281/zenodo.13932057
2) Soares, F. M., Ferreira Pires, L., Olavo Bonino da Silva Santos, L., Corrêa, F. E., de Abreu Moreira, D., Pignatari Drucker, D., Braghetto, K. R., Botazzo Delbem, A. C., & Mauro Saraiva, A. (2025). Agriculture Product Types Ontology (APTO) (v1.31). Zenodo. https://doi.org/10.5281/zenodo.15008549

ACKNOWLEDGEMENTS

F.M.S. acknowledges support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under grant numbers 21/15125-0 and 22/08385-8. K.R.B. received funding from FAPESP (Grant n. 2023/00779-0). F.M.S., D.P.D., K.R.B., F.E.C., D.A.M., A.C.B.D., and A.M.S. thank the Center for Artificial Intelligence (C4AI), a partnership of USP, IBM, and FAPESP (Grant number 2019/07665-4). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001. We also thank the designer Rubens Rangel Silva for his support with this paper figures' preparation. During the preparation of this work the authors used ChatGPT-4o in order to review the text flow, grammar, spelling, and to eliminate redundant text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

[1] R. d. A. Falbo, "Sabio: Systematic approach for building ontologies," Onto. Com/odise@ Fois, vol. 1301, 2014.

[2] N. Casellas, Ontology Evaluation through Usability Measures. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, vol. 5872, pp. 594–603.

[3] B. Albert and T. Tullis, Measuring the user experience: collecting, analyzing, and presenting usability metrics. Newnes, 2013.

[4] B. Fu, N. F. Noy, and M.-A. Storey, "Eye tracking the user experience – an evaluation of ontology visualization techniques," Semantic Web, vol. 8, no. 1, p. 23–41, Nov. 2016.

[5] J. García, F. J. García-Peñalvo, R. Therón Sánchez, and P. Ordóñez de Pablos, "Usability evaluation of a visual modelling tool for owl ontologies," Journal of Universal Computer Science, vol. 17, no. 9, pp. 1299–1313, 2011. [Online]. Available: https://gredos.usal.es/handle/10366/121374

[6] E. García-Barriocanal, M. A. Sicilia, and S. Sánchez-Alonso, "Usability evaluation of ontology editors," KNOWLEDGE ORGANIZATION, vol. 32, no. 1, pp. 1–9, 2005.

[7] Z. H. Malik, "Usability evaluation of ontology engineering tools," in 2017 Computing Conference. London: IEEE, Jul. 2017, p. 576–584.

[8] J. Pak and L. Zhou, A Framework for Ontology Evaluation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 52, p. 10–18.

[9] D. Norman, The design of everyday things: Revised and expanded edition. Basic books, 2013.

[10] E. Schwartz, Exploring experience design: fusing business, tech, and design to shape customer engagement, 1st ed. Erscheinungsort nicht ermittelbar: Packt Publishing, 2017.

[11] J. Nielsen, "How many test users in a usability study?" 2012, accessed: 2024-08-21. [Online]. Available: https://www.nngroup.com/articles/how-many-test-users/

[12] Optimal Workshop, "How to interpret the task results tab in tree testing," 2024, learn how to interpret the different kinds of task scores in tree testing. [Online]. Available: https://support.optimalworkshop.com/en/articles/2626846-how-to-interpret-the-task-results-tab-in-tree-testing

[13] L. A. Esmerino and F. R. Penteado, "Avaliação da qualidade microbiológica da carne de frango comercializada no município de ponta grossa - paraná," Publicatio UEPG: Ciências Biológicas e da Saúde, vol. 17, no. 1, pp. 37–45, jul 2011.

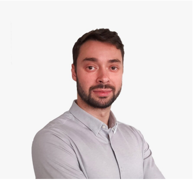
[14] Optimal Workshop, "Interpreting the pietree," 2024. [Online]. Available: https://www.optimalworkshop.com/101guides/tree-testing-101#interpreting-the-pietree

[15] N. Deshmukh, P. S. Rao, H. Sharma, S. K. M.H., L. N. N., and M. K. C.T., "Waste to nutrition: The evolution of whey, a byproduct to galactooligosaccharides production," Food Chemistry Advances, vol. 4, p. 100642, 2024.

[16] G. L. de Paiva Anciens Ramos, J. T. G. aes, T. C. Pimentel, A. G. da Cruz, S. L. Q. de Souza, and S. M. R. Vendramel, "Chapter 19 - whey: generation, recovery, and use of a relevant by-product," in Valorization of Agri-Food Wastes and By-Products, R. Bhat, Ed. Academic Press, 2021, pp. 391–414.

[17] S. M. a. Salgado, N. B. Guerra, and A. B. d. Melo Filho, "Polpa de fruta congelada: efeito do processamento sobre o conteúdo de fibra alimentar," Revista de Nutrição, vol. 12, no. 3, pp. 303–308, Sep 1999.

[18] R. A. Falbo, G. Guizzardi, A. Gangemi, and V. Presutti, "Ontology patterns: clarifying concepts and terminology," in Proceedings of the 4th International Conference on Ontology and Semantic Web Patterns - Volume 1188, ser. WOP'13. Aachen, DEU: CEUR-WS.org, 2013, p. 14–26.



FILIPI MIRANDA SOARES received the Ph.D. degree in Computer Engineering from the University of Sao Paulo, Brazil, in 2025, and is currently a Ph.D. candidate in Computer Science at the University of Twente, the Netherlands, as part of a double degree program, with defense expected for October 2025.

His research focuses on the integration of ontologies, knowledge graphs, and generative artificial intelligence to enhance agricultural data analysis. He currently works as a researcher/project manager in semantic interoperability, ontologies, and data spaces at INRAE (UMR MISTEA, Montpellier, France).

ANTONIO MAURO SARAIVA received the Ph.D. degree in Electrical Engineering. He served as a Full Professor at the Polytechnic School of the University of São Paulo (USP) from 2008 to 2024 and is currently a Senior Full Professor at the Institute of Advanced Studies (IEA) at USP.

He has held various academic leadership positions including Head of the Department, Associate Provost for Research, and Chair of Research Committees at Poli-USP and IEA-USP.

Prof. Saraiva's research interests include agricultural informatics, data governance, and sustainability in knowledge infrastructures.

LUÍS FERREIRA PIRES received the M.Sc. degree from the University of São Paulo in 1989 and the Ph.D. degree from the University of Twente in 1994. He is an Associate Professor in the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente.

His main research interests include service architectures, distributed systems design methodologies, and semantic technologies such as model-driven engineering, ontologies, and linked data.

DEBORA PIGNATARI DRUCKER received the Ph.D. degree in Environment and Society and is a Research Analyst at Embrapa Digital Agriculture in Brazil. She coordinates the GO FAIR Brazil Agriculture Thematic Network and co-chairs the RDA IGAD Community of Practice.

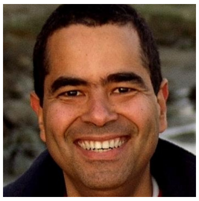
Her work focuses on mobilizing high-quality research data for decision-making in agricultural and environmental policy. She is also involved with the IPBES Data and Knowledge Task Force.

KELLY ROSA BRAGHETTO received the M.Sc. and Ph.D. degrees in Computer Science from the University of São Paulo (USP). She is currently an Assistant Professor at the Institute of Mathematics and Statistics, University of São Paulo.

Her research focuses on data engineering, scientific data management, data integration, and smart city infrastructures.

LUIZ OLAVO BONINO DA SILVA SANTOS received the Ph.D. degree in Computer Science and is an Associate Professor at the University of Twente, Netherlands, and at the BioSemantics group at Leiden University Medical Center.

His research focuses on the FAIR principles, semantic interoperability, ontology-driven conceptual modeling, and intelligent systems design. He is actively involved in international standardization and data stewardship efforts.



DILVAN DE ABREU MOREIRA received the Ph.D. degree in Electronics Engineering from the University of Kent at Canterbury, UK, in 1995, and conducted postdoctoral research in Biomedical Informatics at Stanford University in 2008.

He is currently an Associate Professor at the University of São Paulo (USP), Brazil. His research interests include biomedical informatics, microelectronics, and data science applications in healthcare.

FERNANDO ELIAS CORRÊA received the M.Sc. and Ph.D. degrees in Computer Engineering from the University of São Paulo (USP). He is currently a Postdoctoral Researcher in data management applied to agribusiness at the Center for Artificial Intelligence (C4AI).

His research interests include artificial intelligence, spatiotemporal data processing, and scalable data systems for agricultural innovation.

ALEXANDRE CLÁUDIO BOTAZZO DELBEM is a Full Professor at the Institute of Mathematics and Computer Sciences (ICMC), University of São Paulo (USP). He received his Ph.D. in Computer Science and conducts research in computational intelligence.

His interests include multi-objective optimization, explainable AI, and applications of intelligent systems in health, environment, and engineering domains.