Semantic content negotiation for knowledge exchange in heterogeneous systems

Yousouf Taghzouti[1 [0000−0003−4509−9537]]

Mines Saint-Étienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, F - 42023 Saint-Étienne France
yousouf.taghzouti@emse.fr

Abstract. A resource on the Web can have several representations that we call variants. A user wanting a particular representation of that resource dereferences its Uniform Resource Identifier (URI) with a set of constraints. Content negotiation (CN) is the process by which a server matches the client's constraints to the set of variants the resource has. Typically, media type and language were the predominant dimensions used to differentiate the variants. One can also find the use of the version dimension where variants were created to represent different evolutionary versions of the resource. On the semantic web, a resource can be described using multiple RDF ontologies and adhering to different web ontology language profiles. A finer negotiation of the content is therefore necessary, and the server response must be flexible, especially to answer a request if no representation perfectly validates all the constraints. In this PhD, we aim to address the problem of semantic content negotiation to enable fine-grained negotiation in a heterogeneous web environment. We present our problem statement, including our research questions to address a range of use cases. We also describe our methodology, and provide preliminary results as well as the future steps.

Keywords: Content negotiation · semantic web · constraint.

1 Introduction

Uniform Resource Identifiers (URIs) are the means of identifying resources on the Web, where a resource may have one or more representations. Each of these representations, which we call variants or alternatives as in [5, section 1.3], is a different perspective of that resource, for example an image or a textual representation. To this end, the first proposal for the Web already included a negotiation layer that would allow users to retrieve a representation of a resource corresponding to a set of constraints [2]. This layer between the client and the server is also mentioned as one of the key components of the Web design in the Architecture of the World Wide Web [11, section 3.2].

Negotiation is a communication between two or more parties trying to find a compromise between the needs and constraints of each other. Content negotiation (CN) is the mechanism for choosing the appropriate representation when processing a request. In HTTP one can express and transmit constraints referred to as preferences in [6, section 5.3]. And with that, in addition to finding and transmitting information, it is possible to negotiate more specific dimensions such as media type and languages.

A new dimension of NC was accentuated when handheld devices began to be widely available and used, namely the capacity or characteristics of the device, since different handhelds have different screen sizes and capabilities, e.g., capable of rendering color vs. gray scale. In this use case, content negotiation was also used to select a representation that fit the constraints of the device [3].

For a long time, resources on the Web were primarily intended to be manipulated by humans, and less so by machines, due to a nature that is not semantically-enabled. The inability to understand the content of a representation has impeded their full exploitation by automated content negotiation parties (e.g., a server to select a more appropriate representation). The Resource Description Framework (RDF) [9], the Web Ontology Language (OWL) [8] as well as a variety of vocabularies such as FOAF and Schema.org have enabled the description of Web content and have been the stepping stone to enabling machines to understand it. However, this has raised new CN use cases that are yet to be solved. We mention two of them: one for vocabulary negotiation and the second for RDF shapes.

Negotiation of vocabulary Alice runs a portal that harvests and collects data describing ancient artefacts (creator, material used, date of creation, etc.) from different sources, such as museums and Wikidata. She has noticed that each source uses different vocabularies, sometimes customised. Lately, she has received many requests to find a way to search for data in a specific vocabulary. Or to specify the desired vocabularies in an orderly fashion. For example, exposing creator data using a vocabulary such as FOAF (Friend Of A Friend), Schema.org or DCMI (Dublin Core Metadata Initiative).

Currently, the data graphs available on the APIs use the same media type *text/turtle*. The user has to query all the data graphs manually to select those that use the desired vocabulary.

Negotiation of RDF shapes John is a researcher interested in the evolution of youth unemployment in different societies; he needs data in the form of RDF data graphs. To do this, he queries the data graphs available in various web APIs provided by the university portal. These cases are plausible:

Case 1 - The shape is also important: John needs a representation that conforms to a specific shape. Therefore, vocabulary negotiation is not enough because he would have to manually validate all returned data graphs with the desired vocabularies.

Case 2 - Flexibility vs rigidity: In case 1, the negotiation can be rigid in case John wants *all* the constraints to be valid, and prefers not to have an answer otherwise. Else, the negotiation can be flexible in case John agrees to receive a representation even if it does not satisfy all the constraints.

Case 3 - Not all constraints are equally important: For John, not all shape constraints have the same degree of importance. He therefore wants a way to express this importance for each constraint and to obtain the representation that minimises the violation rate by taking it into account.

The remainder of this paper is structured as follows: Section 2 presents our assumptions, the problem statement, as well as the research questions and the scope of our research, followed by the state of the art and related work in Section 3. Section 4 describes the planned research methodology, while Section 5 briefly describes the proposed approach and preliminary results, followed by the conclusion and future work in Section 6.

2 Problem Statement

In this study, we are interested in how to make CN more refined in/with the semantic web by 1) adding the ability to negotiate a more specific semantic web resource, e.g. the ability to request a data graph that uses certain vocabularies; 2) taking advantage of the semantic web when executing the NC, e.g., requesting that a representation conform to some RDF shapes sent as a constraint in the request. We are also interested in increasing the use of CN in web servers where appropriate. We have these assumptions: A1) Creating a resource that provides an overview of CN characteristics (CN dimensions, CN styles, etc.), in addition to providing CN use cases with their existing or potential solutions, would encourage CN usage; A2) Semantic web technologies can contribute to the CN mechanism; A3) Validation languages, e.g. SHACL, are not only useful for validating RDF graphs, but can be used to introduce flexibility in the process of choosing the best variant. As a first step, we focus on these research questions:

RQ1 What are the characteristics of a CN use case, and how can they be compared and classified?
RQ2 How to formalise the different styles and dimensions of CN?
RQ3 How to use semantic validation to request the best representation that partially validates the request among a set of variants?

Some of the secondary questions that arise are: what are these dimensions? how do we express them, convey them, and how to match them? Dimensions can take many forms, starting with media type, language, and content encoding [6], through constraints on the device's ability to process a certain representation [12], to being more complex to indicate semantic interpretations [1]. Also, for a server, in addition to the interpretation and matching of dimensions, another problem is the explanation of the choice made, either of the selected representation or of the variants provided that were deemed close enough to the one requested.

At this stage in the thesis, we have defined the scope of the research to exclude network conditions, battery status, monetary cost, and other constraints that might alter the overall behavior of the server in an imperfect setting, such as sending a smaller representation size to compensate for poor network bandwidth.

Also, while we are interested in the CN mechanism in general, we will primarily focus on RDF sources [4]. This scope is also reflected in the formalisation, as we started with the document class and then its subclasses, but we did not focus on the speed of the data streams for example. However, it is worth mentioning that this scope did not prevent us from including use cases, such as the use case of negotiating the frequency of updating things (data freshness) in a Web of Things environment when collecting CN use cases to answer [**RQ1**].

3 State of the art

CN has been proposed as an essential layer of the Web architecture since the beginning [2], and has been implemented in the HTTP protocol by providing the means to negotiate variants through headers among others: *accept* to express a constraint on the format of the representation, and *accept-language* to select the preferred language [6]. As well as a set of response status codes to be used to indicate if a specific HTTP request has been completed, but in most cases these codes are generic and do not provide any explanation.

With HTTP, different styles of CN have emerged, including proactive CN, which makes the server responsible for choosing the best variant, and reactive CN, in which the server provides a list of variants and it is up to the client to choose the best one. Also worth mentioning is transparent CN [10], which allows proxies to choose on behalf of the server by taking advantage of the *vary* HTTP header [6]. Unfortunately, the basic HTTP headers are not sufficient, but the protocol can be extended with other headers. Examples of custom headers are *prefer* to request that certain behaviours be used by a server [18], *accept-presentation* to negotiate the RDF presentation [14], *accept-schema* to request how the resource should be structured [19]. This approach solves the problem but is not scalable considering the fact that a new header is created for each new requirement. In addition, interoperability is not achieved since the new header must be known in advance.

Capabilities are another dimension that must be negotiated. For the user agent the *user-agent* header was used but when mobile devices came on the scene a new approach was needed, CC/PP (Composite Capabilities/Preference Profiles) and UAProfile were introduced to solve this problem [12,16]. But even today, as the use of the Web of Things grows, negotiation must take into account their limitations, such as low processor and battery power.

RDF is intended to describe resources on the web using ontologies. And to negotiate these representations, the Dataset Exchange Working Group proposed the profile vocabulary [1], and ways to negotiate profiles that could take the form of constraint language resources [21,20] e.g. SHACL [13] or ShEx [17]. However, this lacks flexibility, for example in the case where a server has to choose between two representations that only partially validate a profile.

To address the performance and privacy impacts of sending headers that are not reliably used in processing a client request, a recent proposal is Client Hints [7]. An *Accept-CH* response header is introduced that servers can use to announce their use of request headers for proactive CN. However, in our opinion, the adoption of headers contributing to CN is insufficient due to the lack of a guidance resource. This resource should document the approaches and features of NC, and present them in a digestible way in the context of use cases. The latter would include their respective or potential solutions and pointers to other links for further investigation.

4 Methodology

The methodology adopted in the development of this doctoral study is as follows:

1. Investigate the state of the art of research relevant to the identified problem. This includes the study of the literature on CN techniques, formulation and validation of constraints in the Semantic Web domain.
2. Develop a Web Resource that collects CN use cases and requirements of these use cases, CN styles, dimensions and constraint conveyance means. This resource aims to supports the categorisation and comparative evaluation of CN features.
3. Initiate the formalisation of the CN problem following a bottom-up approach as described in the Section 5.2.
4. Investigate flexible and explainable CN for RDF resources. The flexible CN techniques will first be evaluated using a synthetic dataset and profiles, followed by real data for future work. Performance and scalability will be measured using response time for a set of predefined graphs and data profiles. Response time will also be measured by filling the data graphs with more and more triples, in order to compare performance with an increasing number of triples (e.g. 500, 1000, 1500 triples). The same measurement technique will be applied for an increasing number of constraints in the profiles by adding shapes in the SHACL case.

5 Preliminary results

5.1 Content Negotiation Theoretical Framework

We believe that CN, as one of the pillars of the Web, deserves to be studied in more depth. We plan to start with a well-designed documentation and presentation for newcomers who want to learn about CN, its features and use cases, or for experts who want to keep up with the latest techniques and technologies used [**RQ1**]. To this end, we propose to create a theoretical framework to categorise and benchmark different use cases of CN, and we want to materialise this in a resource taking the form of a website. CNTF (Content Negotiation Theoretical Framework) is our contribution that categorises the characteristics of CN into different CN groups: style, dimension, etc. CNTF aims to collect use cases of CN, highlight existing solutions if available, or suggest plausible ways to advance them. These are the main objectives of the CNTF. Later, it will be used for the dissemination of our new propositions to advance CN. CNTF attempts to meet these requirements:

Navigable design: CNTF should have a navigable design that includes the ability to move from one concept to another, e.g., from a use case to the CN dimension or style being used via links.

Extensible: One of the main contrasts between a traditional survey document and the CNTF resource is that it should be extensible by allowing for the addition of new CN concepts, e.g., a new dimension, and be up-to-date with different terminology and definitions.

Categorisable: The CNTF must provide the means to categorise different CN use cases and techniques to allow for comparative evaluation. It should also have a well thought out grouping and modelling of different CN concepts to facilitate understanding.

Maintainable: CNTF should promote maintainability by adjusting the used model e.g., community-recommended vocabularies, and by taking into account feedback provided by CNTF users to clarify and rectify the content.

We published a peer-reviewed state of the art at a national conference [22] and the un-reviewed English translation is also made available [23]. CNTF is available at a persistent URI using w3id[1], and a paper describing it was submitted to the resource track of ISWC [24]. Please refer to the *updates* section in CNTF to perceive the list of features already added and those to come.

5.2 A bottom-up formalisation of CN

Another way to study semantic CN is to have more precision and rigour in its formalisation. For this we suggest a bottom-up formalisation of CN [**RQ2**]. Although there have been attempts to formalise CN, it has been done mainly to fit a certain context, such as content adaptation [15]. In our path to formalise CN in semantic contexts, we plan to take a step-by-step approach from basic web document negotiation to specialised graph document negotiation using semantic validation languages e.g. SHACL. The goal is to define in a formal way what a client request with its constraints is, what a server must provide as an answer, what some of the strategies are to choose the representation to return. Then, we progressively add complexity such as how to express a redirection when the server cannot satisfy the request but can provide a URI that would help. And later on, we can define the protocols formally.

5.3 A step towards semantic CN

To solve the use cases already presented and achieve flexible CN in a practical way [**RQ3**], we use SHACL. Specifically, we use the recently introduced header *accept-profile* to request a variant that validates a set of constraints in the form of SHACL documents [20]. A client makes a request with a SHACL document, the server in the traditional procedure has the set of profiles corresponding to the

[1] `https://w3id.org/cntf/`

variants, if a variant conforms to the requested profile it is served otherwise a 406 (Not Acceptable), a 404 (Not found) or a 300 (Multiple Choices) status code are returned depending on the server configuration[2]. In our approach, we propose to validate on the fly the requested profile with the list of available variants, and to provide the closest variant if several partially validate the constraints. We developed a simple algorithm to show how this could be achieved.

The algorithm takes as input a list of SHACL document URIs S that represent the client's constraints, each with an optional numeric value q_s indicating the preference of that profile. The server has a list of data graphs (variants) G from which to choose. The output of this algorithm is a data graph that has the minimum number of violations. For each of the SHACL documents $s \in S$, we validate the available data graphs and record the number of tested constraints n_c as well as the valid constraints v_c. Then, we compute the validation measure with the formula:

$$v_m \leftarrow \frac{v_c}{Max(1, n_c)} \times q_s$$

Once each pair (shape document, data graph) has a validation measure, we deliver the one with the best score. A result worth mentioning is a functional demonstration of CN using profiles that takes the form of SHACL documents. The implementation was done using Java, and Spring Framework to handle queries and intercept query headers. Jena Framework was used to handle RDF graphs and SHACL document and validation[3]. Once we have validated the scalability of the results, we will add the implementation to CNTF.

6 Conclusions and future work

In this paper, we analyse the state of existing work in CN and point to a new direction: the use of validation languages to express finer constraints and achieve semantic CN. We believe that in real-world applications this would make the negotiation process much more flexible. While in the current state, we are only able to negotiate the *text/turtle* media type for example, but we are not able to specify an ontology or shapes that the representation should conform to. We have proposed the CNTF resource, the general formalisation of the negotiation process, and an algorithm for constraint negotiation using SHACL. Our future work will focus on enriching CNTF with more use cases and finalising the features to satisfy the aforementioned requirements, as well as extending the formalisation to include negotiation of specific documents such as RDF graphs using profile negotiation. The results of this study could be used to validate and assert that SHACL introduces flexibility in the best representation selection process.

Acknowledgments I would like to thank my supervisors, Prof. Antoine Zimmermann and Prof. Maxime Lefrançois, for their valuable feedback and support.

[2] Multiple discussions within the Data Exchange Working Group to address these issues, e.g. `https://github.com/w3c/dx-connegp/issues/5`

[3] `https://github.com/YoucTagh/flexible-cn`

References

1. Atkinson, R., Car, N.: The Profiles Vocabulary, W3C Working Group Note 18 December 2019. W3c working group note, W3C (Dec 18 2019)
2. Berners-Lee, T., Cailliau, R., Groff, J., Pollermann, B.: World-Wide Web: The Information Universe. ENRAP **2**(1), 74–82 (1992)
3. Butler, M.H.: Implementing content negotiation using CC/PP and WAP UAProf. HP Laboratories Technical Report HPL (190) (2001)
4. Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1 Concepts and Abstract Syntax. W3c recommendation, W3C (Feb 25 2014)
5. Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P., Berners-Lee, T.: Hypertext Transfer Protocol - HTTP/1.1. RFC 2616, IETF (1999)
6. Fielding, R., Reschke, J.: Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content. RFC 7231, IETF (Jun 2014)
7. Grigorik, I., Weiss, Y.: HTTP Client Hints. RFC 8942, IETF (Feb 2021)
8. Group, W.O.W.: OWL 2 Web Ontology Language Document Overview (Second Edition). W3c recommendation, W3C (Dec 2012)
9. Hayes, P., Patel-Schneider, P.: RDF 1.1 Semantics, W3C Recommendation 25 February 2014. W3c recommendation, W3C (Feb 25 2014)
10. Holtman, K., Mutz, A.: Transparent Content Negotiation in HTTP. Tech. rep., IETF (Mar 1998)
11. Jacobs, I., Walsh, N.: Architecture of the World Wide Web, Volume One, W3C Recommendation 15 December 2004. W3c recommendation, W3C (Dec 15 2004)
12. Klyne, G., Reynolds, F., Woodrow, C., Ohto, H., Hjelm, J., Butler, M., Tran, L.: Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3c recommendation, W3C (Jan 15 2004)
13. Knublauch, H., Kontokostas, D.: Shapes Constraint Language (SHACL), W3C Recommendation 20 July 2017. W3c recommendation, W3C (Jul 20 2017)
14. Lefrançois, M.: RDF presentation and correct content conveyance for legacy services and the web of things. In: Proc, 8th International Conference on the Internet of Things. pp. 43:1–43:8. ACM Press (Oct 2018)
15. Lerouge, S.: Personalizing quality aspects for video communication in constrained heterogeneous environments. Ph.D. thesis, Ghent University (2006)
16. Open Mobile Alliance: User Agent Profile. Tech. rep., Open Mobile Alliance (May 2003)
17. Prud'hommeaux, E., Boneva, I., Labra Gayo, J., Kellogg, G.: Shape Expressions Language 2.1. W3c community group report, W3C (2019)
18. Snell, J.: Prefer Header for HTTP. RFC 7240, IETF (2014)
19. Svensson, L.: An http Header for Metadata Schema Negotiation. In: W3C Workshop on Smart Descriptions & Smarter Vocabularies (SDSVoc). W3C (Nov 2016)
20. Svensson, L., Atkinson, R., Car, N.: Content Negotiation by Profile, W3C Working Draft 26 November 2019. W3C Working Draft, W3C (Nov 26 2019)
21. Svensson, L., Verborgh, R., Van de Sompel, H.: Indicating, Discovering, Negotiating, and Writing Profiled Representations. Internet draft, IETF (Mar 2021)
22. Taghzouti, Y., Zimmermann, A., Lefrançois, M.: Négociation de contenu sur le web: un état de l'art. In: Journées Francophones d'Ingénierie des Connaissances (IC 2022) @ Plate-Forme Intelligence Artificielle (PFIA 2022) (2022), to appear
23. Taghzouti, Y., Zimmermann, A., Lefrançois, M.: Content negotiation on the Web: State of the art. arXiv (2022)
24. Taghzouti, Y., Zimmermann, A., Lefrançois, M.: A portal for the state of the art on content negotiation. In: ISWC Resource Track. Springer (2022), under review