

Conceptual Modeling of Explainable Recommender Systems: An Ontological Formalization to Guide Their Design and Development

Marta Caro-Martínez

MARTCARO@UCM.ES

Guillermo Jiménez-Díaz

GJIMENEZ@UCM.ES

Juan A. Recio-García

JARECIOG@UCM.ES

Department of Software Engineering and Artificial Intelligence

Universidad Complutense de Madrid

Madrid, Spain

Abstract

With the increasing importance of e-commerce and the immense variety of products, users need help to decide which ones are the most interesting to them. This is one of the main goals of recommender systems. However, users' trust may be compromised if they do not understand how or why the recommendation was achieved. Here, explanations are essential to improve user confidence in recommender systems and to make the recommendation useful.

Providing explanation capabilities into recommender systems is not an easy task as their success depends on several aspects such as the explanation's goal, the user's expectation, the knowledge available, or the presentation method. Therefore, this work proposes a conceptual model to alleviate this problem by defining the requirements of explanations for recommender systems. Our goal is to provide a model that guides the development of effective explanations for recommender systems as they are correctly designed and suited to the user's needs. Although earlier explanation taxonomies sustain this work, our model includes new concepts not considered in previous works. Moreover, we make a novel contribution regarding the formalization of this model as an ontology that can be integrated into the development of proper explanations for recommender systems.

1. Introduction

Nowadays, a huge amount of information can be found on the Internet. With the appearance of online shops, like Amazon or eBay, and platforms for entertainment consumption, like Spotify or Youtube, internet users are able to obtain and consume products easily and rapidly. However, the amount of products offered by these platforms is immense, and the sheer quantity can hinder the task of finding the product that the user really wants. Recommender systems alleviate this problem by helping users to find products that are interesting to them. Thus, users make better decisions thanks to the recommendations provided by those systems.

Many times, users do not trust recommender systems as they do not know how the recommendations are carried out or the reasons why a product is recommended. This problem causes a poor user experience that may lead to a failure of the recommender system. Therefore, explanations are necessary to help users understand the reasons behind a recommendation. These explanations may serve many different purposes. They can show how the

system works (transparency) or help users make an informed choice (effectiveness). They may be evaluated on whether they convince the user to make a purchase (persuasiveness) or whether they help the user make a decision quickly (efficiency). Explanations should consider different user profiles with custom needs (target users, stakeholders, developers,...). It is also critical of the way of presenting recommendations and how users interact with the provided explanations. The knowledge available in the recommender system and the algorithms applied to generate the recommendations restrict these previous aspects.

Consequently, *successful* explanations must be carefully designed, considering all these features, their interrelations and dependencies. Our work aims to alleviate this problem by introducing a new conceptual model for explanation systems to achieve a successful explanation design. A *conceptual model* is a set of concepts (entities and features) organized considering their relationships to define a more general concept with the goal of guiding research on this concept (Green, 2014). Our approach provides a refinement concerning previous models that are analyzed and integrated. This way, we provide an integrating framework that enhances existing models with novel concepts that relate them and define conceptual dependencies.

The model proposed in this paper allows the design of explanations for recommender systems following ontology-guided development methodologies. In our previous research, we have validated such approaches to generate recommender systems according to the actors (researchers, developers, etc.), their backgrounds, motivations and goals (Recio-García, González-Calero, & Díaz-Agudo, 2014b; Jorro-Aragoneses, Ceron-Rios, Díaz-Agudo, Recio-García, & López-Gutierrez, 2017). This development process relies on semantic descriptions of work-flows and components using an ontology, RecOnto¹, which supports the reasoning regarding the correctness of the recommender system being generated. The work presented in this paper continues this research line, describing our proposal to formalize a model that complements RecOnto to include the semantic description of explanations for recommender systems.

To demonstrate the suitability of the proposed model and its associated ontology, we present an exhaustive validation by classifying reference explanation systems for recommender systems found in the literature. This way, we show that our approach can categorize not only the explanation capabilities of recommender systems but also explanations generated by other types of intelligent systems, such as machine learning or scheduling applications. An early version of this conceptual model was presented in our previous work (Caro-Martinez, Jimenez-Diaz, & Recio-Garcia, 2018) at XCBR workshop. However, our current work extends the model and the validation process, as long as it defines the ontology that can be employed on an ontology-guided development methodology.

The paper is structured as follows. Section 2 reviews state-of-the-art literature about explanations in recommender systems. In Section 3, we relate the basis of the model's representation. Next, Section 4 describes in detail our model and Section 5 relates the model validation process. Section 6 introduces the ontology implemented and, finally, Section 7 concludes the paper.

1. An interactive visualization of the RecOnto ontology is available at <https://gaia.fdi.ucm.es/ontologies/#reconto>

2. Literature Review

In this work, we have studied several publications where different ways of explaining recommendations can be found. A number of them are studies that propose ways of classifying explanations (Herlocker, Konstan, & Riedl, 2000; Nunes & Jannach, 2017; Gedikli, Jannach, & Ge, 2014; Tintarev & Masthoff, 2015; Papadimitriou, Symeonidis, & Manolopoulos, 2012; Tintarev & Masthoff, 2012; Eiband, Schneider, & Buschek, 2018; Friedrich & Zanker, 2011; He, Parra, & Verbert, 2016; Al-Taie & Kadry, 2014). Moreover, we have studied different explanation systems in order to validate and enhance the categories extracted from the previous surveys. In the following subsections, we will review these works.

2.1 Surveys and Classification Models

Initially, we studied some works about XAI (eXplainable Artificial Intelligence) to get a broad vision about explanation systems in a wide variety of artificial intelligence approaches. For example, we delved into the wide study stated by Miller (2019), where the author describes the importance of explaining artificial intelligence from a social point of view. It claims that the process of an explanation generation is divided into three steps: 1) the *cognitive process*, where we find out the causes (called *explanans*) behind an event (the *explanandum*); 2) the creation of the *product* (the resultant explanation) according to the causes identified before; and 3) the *social process*, where the explainer gives the knowledge to understand the event to the explainee. We also reviewed the work by Adadi and Berrada (2018), where the authors do a review about XAI as a current key factor in the artificial intelligence field. It describes many aspects of explanations: the concept of XAI, its applications and domains, related surveys and taxonomies. It mainly focuses on research trends. Another work that we have examined is the survey by Das and Rad (2020). In this case, the review focuses on the explanation scope, differences among the methodologies and implementations. We can also find an interesting taxonomy is the proposal by Arrieta et al. (2020), where the authors define explainability in machine learning making an in-depth review on XAI. They provide a set of useful concepts to understand the type of existing explanations in this type of artificial intelligence techniques.

Later, we began to research on studies specifically related to explanations in recommender systems. First, we studied the survey proposed by Tintarev and Masthoff (2015), where we found an analysis of the different purposes of explanations in recommender systems. The survey describes the criteria that *good* explanations should have, the different ways of presenting recommendations and how users interact with recommender systems. The work by Gedikli et al. (2014), which is a mix between survey and explanation proposal, we also found the properties described by Tintarev and Masthoff (2015), but this knowledge was applied to different ways of presenting the explanations for recommendations. The study proposed by Al-Taie and Kadry (2014) also makes a review of different visualization approaches that have been applied in recommender systems found in the literature. Thanks to these studies and the proposal by Herlocker et al. (2000), we discovered new ways of visualizing explanations that we will consider in our model.

The work by Eiband et al. (2018) compares two models of explanation: a normative model and a pragmatic model. The normative model is focused on transparency and user

understanding, while the pragmatic model is designed for usability and the efficient use of explanations. This approach will help us to refine the features that our model should have.

Additionally, we studied the explanation taxonomies proposed by Nunes and Jannach (2017), Papadimitriou et al. (2012) and Friedrich and Zanker (2011). The most detailed work is the systematic review detailed by Nunes and Jannach (2017). It describes many aspects considered in the previous studies but improves them with new categories. It also presents detailed descriptions of different explanation types according to these aspects. The works by Papadimitriou et al. (2012) and Friedrich and Zanker (2011) incorporate different innovative ideas about explanations and their classification. Papadimitriou et al. (2012) present new concerns about explanations in social recommender systems, and Friedrich and Zanker (2011) introduce new facets regarding recommendation types and the importance of understanding these types in order to explain recommendations.

In the work by Tintarev and Masthoff (2012), which provides a study on the effectiveness of explanatory information and how it is evaluated, we can also find some interesting details about recommender systems and the types of knowledge sources for explaining recommendations. It will help us to complete our ideas about explanation goals and types of knowledge sources. Moreover, Zhang and Chen (2018) did a helpful review where we encounter a classification of explanations regarding their knowledge source, such as feature-based or opinion-based explanations, and their type of visualization, such as textual or visual explanations. We also studied the work by He et al. (2016), where we found a survey about visualization of explanations, emphasizing graphic visualization such as diagrams, plots or trees. It also includes a classification of these explanations according to their objectives. In relation with this previous work, we can observe examples of systems that show recommendations in an innovative visual mode using novel diagrams and graphs, like the ones described by Verbert, Parra, Brusilovsky, and Duval (2013) and Parra, Brusilovsky, and Trattner (2014).

2.2 Explanation Approaches

After studying survey and review works, we extended our revision to concrete explanation approaches for recommender systems, where we also found publications with relevant contributions. One of the most exhaustive works is the approach proposed by Quijano-Sanchez, Sauer, Recio-Garcia, and Diaz-Agudo (2017). Besides the system proposed, this work delves into the study of explanations in recommender systems. From both the study and the explanation approach proposed, we extracted ideas about explanation goals, explanations based on social information, groups as target users, or different visualization formats.

Sharma and Cosley (2013) introduce several types of social explanations in a music artist recommender system. Namely, this work promotes the use of social relationships as a way of explaining recommendations. In the works by Musto, Narducci, Lops, De Gemmis, and Semeraro (2016), and Sansonetti, Gasparetti, Micarelli, Cena, and Gena (2019), we find out other kinds of knowledge sources to implement explanations, like Linked Open Data (LOD). The proposals by Lu, Dong, and Smyth (2018) and Peña, O'Reilly-Morgan, Tragos, Hurley, Duriakova, Smyth, and Lawlor (2020) describe innovative models to mix textual information from user reviews and knowledge from matrix factorization to make explanations. These four works noticed us a wide variety of knowledge sources used in explanation systems.

The publication by Lamche, Adgüzel, and Wörndl (2014) is a detailed description of an interactive explanation for recommender systems in a mobile application. In particular, the most useful ideas found here are the use of interactive explanations, the way of presenting an explanation in mobile applications and the analysis of the different perspectives that an explanation can have. The work by Buchanan and Shortliffe (1984) proposes a rule-based explanation system. They describe an example of explanations that justify the system’s behavior using its own internal working process. This work led us to distinguish between introspective or post hoc explanation approaches, described later. Moreover, it depicts some secondary goals for the explanations being generated, like education and debugging. Berkovsky, Taib, and Conway (2017) made a study that analyzes the significance and impact of trust in the success of recommender systems. Specifically, it focuses on comparing different explanation dimensions, presentations and priorities. The work by Zanker and Ninaus (2010) propose a knowledge-based framework for the generation of explanations. The goal of this publication is to provide transparency for recommendations based on the coincidences between item features and user preferences. From this study, we extracted the idea of classifying explanations according to the knowledge source used to generate them. In the recent work by Tsai and Brusilovsky (2020), authors present a hybrid recommender system that allows users to watch explanations and control them. We confirmed some of the ideas we had already collected with this work, as the importance of the knowledge source, types of visualizations, and critical properties of explanations, like scrutability or transparency.

3. Model Representation

One of this paper’s major contributions is the formalization of a conceptual model of explanations through the semantic constructions and vocabulary used to define ontologies (Euzenat, Shvaiko, et al., 2007). In any semantic model, its concepts must be related using meaningful relationships. Concretely, we use the basic relationships found in any ontology representation language: *composition*, *subconcept* and *instance*. A composition is established when some concepts of our ontology are a combination of other concepts that define them from different perspectives. However, a subconcept is an extension of a concept that refines the definition of this concept, so it follows an “*is-a*” relationship. Subconcepts are not disjoint so that explanations can be classified into several concepts. Lastly, an instance is a concrete instantiation, example or case of a concept. This way, we solve the lack of formalization found in the previous studies and will be able to define the ontology presented in Section 6. To do so, we need to describe the background knowledge on recommender systems that supports our model’s definition.

3.1 Domain Entities About Recommender Systems

Our work focuses on building a model that should help to design effective explanations for recommender systems. In order to facilitate the understanding of the model proposed here, it is necessary to define some important concepts related to recommender systems. The main entities involved in recommender systems are *users* and *items*. *Users* are the people interested in getting new items and who interact with the system to explore the products

in the recommendation platform. *Items* are the products available in the recommendation system: food, clothes, hotels, experiences, etc.

The recommendation process relies on different sources of data about users and items. On the one hand, the interactions carried out by users in the recommendation system are one of the most exploited sources to recommend new items to these users. One widely used way to identify user preferences is *ratings*, a form of explicit user feedback about an item. If one user rates an item with a good score, then this item (and its features) will be suitable for this user. Moreover, implicit feedback, which represents whether the user has interacted with an item or not, is another extended source to infer user preferences. On the other hand, item descriptions are another kind of data involved in the recommendation process. Every item has its own *content*, an extensive representation of the item; for example, a text description. This content can be represented by a set of *features* that defines the item: its price, color, or size.

Frequently, the recommendation is supported by the *similarities* among users or items. The similarities are the likeness between two elements; for example, between item features or user ratings. Although it is common to recommend items similar to the ones that the user interacted with previously, some recommendation approaches suggest items on the basis of user *preferences*, the requirements elicited by the user before interacting with the recommender system.

Finally, every user is part of the *community* defined by the recommender system. The behavior of both every individual user and the community can be employed in the recommendation process. In addition, the recommendation can be enhanced by using the information extracted from the environment and the circumstances of the user and her interaction with third-party systems. We call this *contextual information*.

Next, the following section details the concrete concepts of our proposed model.

4. Model Description

Thanks to the study carried out, we propose an enhanced model to describe explanations for recommendation systems. Our model introduces four main definition facets: *motivation*, *knowledge*, *generation* and *presentation*. Figure 1 shows an overview of the conceptual model developed in our work that will be detailed in following subsections.

4.1 Motivation

Motivation refers to the goal of the explanation depending on who is target user. Both ideas, goal and user compose the motivation of an explanation.

4.1.1 USER

Recommender systems are supposed to satisfy user needs. However, explanations do not have to be aimed exclusively at the final user. We can classify them based on the user the explanation is addressed to:

Target User. These are the users who receive the recommendations made by the system. Explanations are relevant to the target users because of the need to understand how or why an item is recommended in order to assess if they are interested in it. This is the

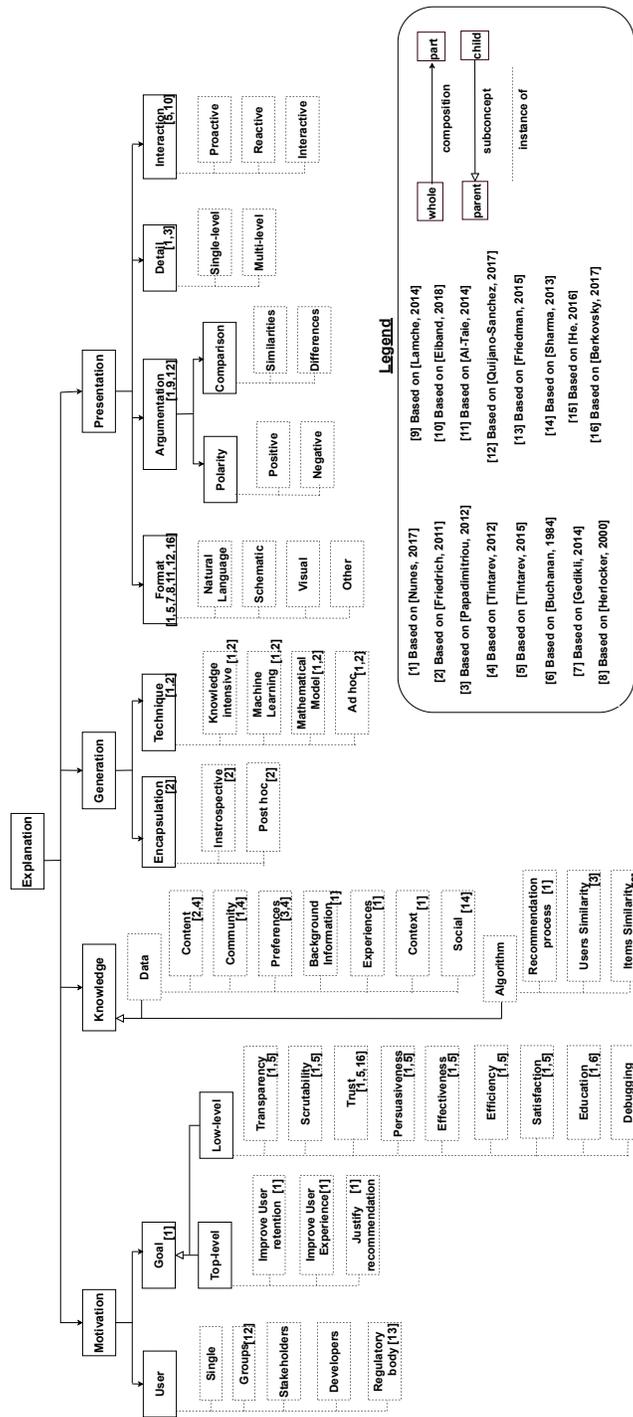


Figure 1: Conceptual Model for explanations in recommender systems.

most common type of user. We can consider two subtypes of target users: single users and groups of users. The target is a single user when the user that receives the recommendation

also receives the explanation. In this case, the explanation is designed in order to justify the recommendation for that concrete user. However, explanations can also be targeted to groups (Jameson & Smyth, 2007; Recio-Garcia, Jimenez-Diaz, Sanchez-Ruiz, & Diaz-Agudo, 2009) and they usually try to describe each member's constraints and conditions that were taken into account when generating the recommendation.

Stakeholders. These are the people interested in the recommender system's success, like product owners, investors or entrepreneurs. They do not receive the recommendation, but they need explanations to measure the recommendation process's performance and success. For example, the IMVEX rule system delivers some explanations to the target user and other ones to the stakeholders, depending on what they need to know (Finch, 1998).

Developer. Recommender system developers need to know how the recommender works to perform different tasks, such as debugging, maintenance or extension. In this case, explanations are focused on helping developers to understand the internal behavior of the system (Robillard, Walker, & Zimmermann, 2010).

Regulatory Body. This represents the legislative and government organizations that need to know how recommender systems work in order to regulate their transparency and correct use (Friedman, Knijnenburg, Vanhecke, Martens, & Berkovsky, 2015). As recommender systems are being widely used nowadays, they are being incorporated into governmental processes. Therefore, a demand increasingly heard from society is that regulatory bodies guarantee their correct behavior. A clear example is the controversy created by the revelation of possible bias in a widely used system for predicting recidivism (Angwin, Larson, Mattu, & Kirchner, 2016). Another example is the EU General Data Protection Regulation (GDPR) which regulates the protection of EU citizens and also envision the right of an explanation in decision-making systems (Wachter, Mittelstadt, & Russell, 2017).

4.1.2 GOAL

The goal of an explanation is closely related to the user to whom the explanation is targeted. The goals can be divided into two subconcepts: top-level goals and low-level goals. Top-level goals are focused on the user, while the low-level ones help to achieve top-level goals.

Top-level goals are divided into three groups, inspired by Nunes and Jannach's proposal (Nunes & Jannach, 2017):

Improve user retention. The goal of the explanation is to increase users' probabilities to return to the recommendation system. This goal is usually very relevant for stakeholder users as they aim to enhance user retention in order to increase profits.

Improve user experience. The main goal of the explanation is to help users make good decisions and enjoy the recommendation activity. This goal is mainly aimed at the target user and the developers in charge of implementing the system. It is also related to the previous goal, as retention is closely influenced by user experience.

Justify recommendation. The goal of this type of explanation is to support the recommendation provided to the user, helping her to understand why an item was recommended.

This type of explanations can help to achieve the previous goals. It focuses on satisfying target users' needs and developers' tasks as system introspection.

Although these are the goals proposed by Nunes and Jannach (2017), other works such as (Tintarev & Masthoff, 2015) and (Buchanan & Shortliffe, 1984) propose more detailed criteria for designing good explanations. We have considered these criteria as low-level goals. We can design an explanation while taking into account one or more of these criteria, which are also related to each other. Following the model suggested by Nunes and Jannach (2017), Tintarev and Masthoff (2015), Buchanan and Shortliffe (1984), we have included the following low-level goals in our model:

Effectiveness. The explanation has to help the user find the items that they need or wants (Tintarev & Masthoff, 2015).

Efficiency. A good explanation helps the user to make decisions quickly (Tintarev & Masthoff, 2015).

Trust. Explanations should increase user confidence in the system. If users understand the explanation and how a recommendation is made, their reliance on the system will be improved (Tintarev & Masthoff, 2015; Buchanan & Shortliffe, 1984; Berkovsky et al., 2017).

Scrutability. When an explanation is scrutible, the user will be able to provide feedback about how the recommender system is working (Tintarev & Masthoff, 2015).

Persuasiveness. An explanation is persuasive if it convinces the user that the recommendation is appropriate (Tintarev & Masthoff, 2015; Buchanan & Shortliffe, 1984).

Satisfaction. Explanations help to increase the user's perceived quality of the recommendations, enhancing user experience, general user enjoyment and pleasure with the system (Tintarev & Masthoff, 2015).

Transparency. An explanation is transparent when it specifies how and why a recommendation is made (Tintarev & Masthoff, 2015; Buchanan & Shortliffe, 1984). Somehow, transparency and justification might seem equals, but they are not. Transparency brings information about how the recommender model creates recommendations. However, justification brings reasons for a recommendation without considering the process carried out by the recommender model.

Education. With explanations, users can learn something that could help them. If users feel that they learn something, the probabilities of a user returning to the recommendation system improve (Buchanan & Shortliffe, 1984).

Debugging. Explanations allow users to identify bugs in the recommendation system (Buchanan & Shortliffe, 1984). Here we also include any other software extension or maintenance task.

4.2 Knowledge

Explanations for recommender systems can be generated based on the type of knowledge used to create the recommendation (Richter & Weber, 2016; Zanker & Ninaus, 2010). Rec-

ommender systems need a specific type and amount of knowledge to achieve interesting items to recommend for a target user. Therefore, we need to consider the source of the knowledge used to generate a recommendation when we design or analyze its related explanation. Thus, we introduced the knowledge as a novel factor in our model. Regarding the existing literature, explanations can extract information from two types of sources: the data that feeds the recommender or the algorithm itself used to generate the recommendation. Based on this observation, we propose two different subconcepts to define the nature of the knowledge used to generate the explanation. In the following sections, we detail these subconcepts, and Table 1 will show examples that illustrate the different types of explanations depending on the knowledge concept.

4.2.1 DATA

These are the explanations based on the information available during the recommendation process. We have identified different types of data used in explanations:

Content. The explanation for a recommended item refers to the item description. It shows non-personalized information, but it helps the user to make good decisions (Friedrich & Zanker, 2011; Tintarev & Masthoff, 2012). For example, “This shirt is white, and its price is 8.99 €.”

Preferences. The explanation refers to the item features while considering the preferences of the user (Papadimitriou et al., 2012; Tintarev & Masthoff, 2012). For example, “This shirt is white, a color that you like, and its price is 8.99 €, less than 10 €, your budget.”

Community. The information used to explain the recommendation is extracted from the behavior of the community of users interacting with the recommender system (Nunes & Jannach, 2017; Tintarev & Masthoff, 2012). For instance, “This shirt is one of the shop’s top sellers”.

Background Information. The data is extracted from the domain (Nunes & Jannach, 2017), such as information from reviews or external interactions. In the work by Muhammad, Lawlor, Rafter, and Smyth (2015), we find a system which explains recommendations by extracting hotel features from the users’ reviews and building a case base from them. The system builds an explanation for the current recommendation based on the knowledge obtained from previous reviews by other users. For instance, “This shirt is made with organic cotton, which is grown using methods and materials that have a low impact on the environment”.

Experiences. This type includes all the explanations that refer to past user experiences and system history (Nunes & Jannach, 2017). For instance, “This shirt is white, and you bought white shirts yesterday”.

Context. This data refers to the dynamic features that describe the requirements of target users. For example, a recommender system for tourist itineraries should adapt the proposal depending on the concrete weather restrictions when the user queries the system. Here, the preferences of the user do not change (they are static knowledge) but the context must be taken into account to adapt the recommendation to the concrete user environment (which

Knowledge	Example
Content	“This shirt is white, and its price is 8.99€.”
Preferences	“This shirt is white, a color that you like, and its price is 8.99€, less than 10€, your budget.”
Community	“This shirt is one of the shop’s top sellers.”
Background Information	“This shirt is made with organic cotton, which is grown using methods and materials that have a low impact on the environment.”
Experiences	“This shirt is white and you bought white shirts yesterday.”
Context	“This shirt is good for warm days like today.”
Social	“Your brother bought this shirt” “Your best friend also bought this shirt.”
Recommendation Process	“This shirt was recommended because the system has estimated that it is the cheapest one that fits your preferences.”
User Similarity	“This shirt was recommended because users similar to you bought it.”
Item Similarity	“This shirt was recommended because you bought similar shirts before.”

Table 1: Examples of textual explanations depending on the knowledge used to generate the recommendation.

is innately dynamic) (Adomavicius & Tuzhilin, 2011). For instance, “This shirt is good for warm days like today”.

Social. The explanation refers to the user’s relationships to justify recommendations (Sharma & Cosley, 2013). For example, “Your brother bought this shirt” or “Your best friend also bought this shirt”.

4.2.2 ALGORITHM

Explanations are based on the information that the system collects from the recommendation algorithm. Explanations can be based on the whole recommendation process, or on a relevant sub-component, mainly user similarity or item similarity. The classes proposed for this kind of knowledge are the following:

Recommendation Process. This type of explanation shows how the algorithm performs a recommendation (Nunes & Jannach, 2017; Eiband et al., 2018). For example, MYCIN is an expert system that selects the minimum number of drugs to cure infections. CHRONICLER explains the MYCIN recommendation by analyzing the algorithm’s decision process, showing the user a chronicle of the system events (Buchanan, 1984). For example, “This shirt was recommended because the system has estimated that it is the cheapest one that fits your preferences”.

User Similarity. The explanation justifies the recommendation through the interactions that similar users have carried out with the items recommended (Papadimitriou et al., 2012). For example, “This shirt was recommended because similar users to you bought it”.

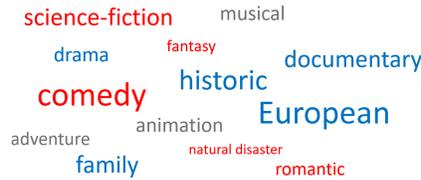


Figure 2: Example of a post hoc graphic explanation: a tag cloud

Item Similarity. Explanations are based on the similarity that exists between items recommended and items that the user interacted with (Papadimitriou et al., 2012). For instance, “This shirt was recommended because you bought similar shirts before”.

4.3 Generation

Explanations for recommender systems can be implemented using different processes and techniques. We define it as a composition from encapsulation and technique.

4.3.1 ENCAPSULATION

During the analysis of these techniques, we observed that recommender systems have two different approaches of encapsulation: as a white box system or as a black box system. A white box system allows users to learn the way its reasoning algorithm operates. On the other hand, a black box system does not provide any knowledge of how it works internally (Friedrich & Zanker, 2011). This way, we can classify explanations according to the encapsulation facet, as *introspective* or *post hoc*, if the recommendation system acts as a white box or black box system, respectively.

Introspective. Explanations are generated by the recommendation algorithm itself because it operates as a white box system (Friedrich & Zanker, 2011). For example, explanations proposed by Herlocker et al. (2000), like histograms with grouping, are introspective. Histograms with grouping is a type of explanation that shows a histogram where users can see similar users’ ratings for the item recommended. Therefore, users know how the recommendation was made –in this case, using user similarity.

Post hoc. The recommender system acts as a black box system, so the explanation system has to implement an external technique for creating justifications on top of the underlying recommendation algorithm (Friedrich & Zanker, 2011). For instance, *tagsplanations*, proposed by Vig, Sen, and Riedl (2009), can be considered as an post hoc explanation system. *Tagsplanations* are tags that show the degree to which the tag describes the item and the user’s preference for the item. The explanation is not based on the recommendation process but is based on content and preferences. In Figure 2, we show a tag cloud based on tagsplanations.

4.3.2 TECHNIQUE

No matter whether the system operates as black box or as white box, different techniques can be used to generate the explanations: case-based reasoning, rules, data-mining, etc. These techniques that generate the explanation can be integrated into the recommendation process itself (white box) or as a post hoc explanation module (black box). Therefore, explanation systems must also be classified according to the algorithm used for generating explanations regardless of their introspective or post hoc nature. The main types of explanation, according to their explanation algorithms, are (Nunes & Jannach, 2017):

Knowledge-based Explanation. These are based on techniques that use a knowledge source to solve problems. Examples of explanation systems based on knowledge are CARES or DIRAS, which use Case-Based Reasoning to explain medical diagnoses (Doyle, Tsymbal, & Cunningham, 2003).

Machine Learning Explanation. According to Mohri *et al.*'s definition (Mohri, Ros-tamizadeh, & Talwalkar, 2012), machine learning algorithms are the computational techniques that use information about experiences to make accurate predictions and better performance. Experiences are defined as past information which the system can access. In this case, these past experiences are used to explain recommendations. For example, the work described by Donkers, Loepp, and Ziegler (2018) applies deep learning techniques to user textual reviews with the purpose of retrieving information and using it to explain recommendations.

Mathematical Model Explanation. A mathematical model can be used to make predictions by using mathematical theories and concepts. Explanations to justify the recommendation can also be implemented through these mathematical models. For example, (Wang, Fi, Liu, & Song, 2016) describes an approach based on a probabilistic model to explain recommendations by using sentiment analysis on reviews.

Ad-hoc Explanation. The explanation algorithm is created *ad hoc* to provide the explanations required by the recommender system. For example, by Barbieri, Bonchi, and Manco (2014), link prediction techniques, which are techniques of social network analysis, are used to explain recommendations in a system that represents users' interactions as a graph.

4.4 Presentation

The presentation concept represents the way the system displays an explanation to the user. Among all of the publications studied, we have identified four different main sub-concepts related to presentation: the display format, the argumentation, the level of detail and the interaction approach.

4.4.1 DISPLAY FORMAT

Explanations can be shown in different ways, more or less elaborate. The following ones are the main formats that we have found in the literature (Herlocker et al., 2000; Gedikli et al., 2014; Nunes & Jannach, 2017; Tintarev & Masthoff, 2015; Al-Taie & Kadry, 2014; Quijano-Sanchez et al., 2017):

This movie was recommended to you because:

- It is a **comedy**, which is your favorite genre.
- The main actress/actor is **Hugh Grant**, whom you like.
- The movie was filmed in the **90's**. Most of your favorite movies were filmed in this decade.

Figure 3: Example of an explanation in natural language

Rating	Number of similar users
★	3
★★	2
★★★	2
★★★★	10
★★★★★	36

Similar users' ratings for the movie recommended

Figure 4: Example of a schematic explanation

Natural Language. The justifications for a recommendation are displayed as a textual description (Figure 3). In this case, it is common to employ templates where a predefined text is shown for every recommendation while modifying the values associated with that concrete recommendation. This is the most common method for presenting an explanation. We can observe a natural language explanation in the CBR system CARES (Doyle et al., 2003), which shows why a complication can appear in diabetes patients.

Schematic. Explanations are shown not in natural language but in a simplified text format like tables, lists, tags, logs or ratings. An example of a schematic explanation found in the literature is the Herlocker *et al.* proposal for MovieLens, which justifies why a movie is recommended to a user with ratings from her neighbors (Figure 4).

Visual. Explanations show justifications through a more graphic way, such as charts like histograms or tag clouds. One example of tag clouds is the Gedikli *et al.* proposal, which defines an item with descriptive tags. The tag color represents the user's sentiment with regard to that feature (Figure 2).

Other. Explanations are generated using more innovative display formats; for example, audio or video. However, there are not many studies that include these categories. Nunes and Jannach's taxonomy (Nunes & Jannach, 2017) does include an example of a system that uses audio and voice explanations to advise non-expert users about what electric devices they should use (Terano, Suzuki, Onoda, Uenishi, & Matsuura, 1989). We have also found explanations developed in innovative technologies, like augmented reality, for recommender systems, in the field of tourism (Gavalas, Konstantopoulos, Mastakas, & Pantziou, 2014).

4.4.2 ARGUMENTATION

Explanations can show different outlooks for a recommended item: they can present the positive arguments, or they can disclose the negative arguments of this item for the user (Nunes & Jannach, 2017; Quijano-Sanchez et al., 2017; Lamche et al., 2014). Additionally, these arguments can be supported by using the similarities or differences between item features and user preferences.

4.4.2.1 Polarity Regarding polarity, we can define an explanation as:

Positive. This type of explanation justifies a recommendation with positive arguments. These explanations tell the user why a recommended item is suitable for her. For instance: “This shirt is medium size, which is the feature that you specified”.

Negative. Sometimes, explanations show features that do not suit user preferences and tastes by using negative argumentation. For example, “This shirt is green, the color that you are looking for. However, it is made of polyester and you want cotton shirts”.

4.4.2.2 Comparison The comparison concept refers to the argumentation base used to justify a recommendation.

Similarities. These explanations highlight the similarities between item features and user preferences. This type is positive when the explanation highlights the correspondence between user likes and item features. However, it is negative if the user wants to discard similar items because she is looking for novelty or diversity. For example: “This shirt is green, your favorite color” could be positive if the user always wants the same type of items, or negative if she wants items with new features.

Differences. These explanations emphasize the differences between the items recommended and the user’s interests. An example is: “This shirt is yellow, which is not your favorite color, but you might also like it”. Although this type of explanations can be perceived as negative, it can also be positive when the explanation needs to justify the diversity (Ruiz-Iniesta, Jimenez-Diaz, & Gomez-Albarran, 2009) or the serendipity (Lamche et al., 2014) of the recommendation.

4.4.3 DETAIL LEVEL

The level of detail is the degree of specification about a recommendation included in its explanation. The level of detail may change for different explanations, depending on what the user needs to know or what the developer wants to show. We consider two types: single-level and multi-level, following the classifications by Nunes and Jannach (2017), Papadimitriou et al. (2012). Due to the level of detail being subjective, we define this concept by taking into account three features: the knowledge base, the number of visualization types and the length of the explanation:

Single-level. Explanations are briefly detailed. Based on previous aspects in the model, we define a single-level explanation as one with a single knowledge base and has a single way of showing the explanation and/or is a short explanation with few details. For example, the explanations proposed by Herlocker et al. (2000) are considered single-level by our model

because they show a single type of visualization and very few details. A graphical example of this kind of explanations is shown in Figure 4.

Multi-level. On the other hand, a multi-level explanation uses several knowledge sources, has different visualization formats and/or it is a long explanation with many details. For example, the tag cloud presented by Gedikli et al. (2014) is a multi-level explanation because it uses two different knowledge sources based on content and preferences.

4.4.4 INTERACTION

The different interaction approaches depend on how the user obtains explanations through the system (Tintarev & Masthoff, 2015; Eiband et al., 2018). Interaction approaches can be classified in the following sub-concepts:

Proactive. The system is in charge of initiating the explanation. It is provided with the recommendation, and the user does not have to ask for a justification because the explanation is shown by default. We find examples of proactive explanations in commercial platforms like Amazon, eBay or Netflix.

Reactive. In this case, the user asks for an explanation of the recommendation received. For instance, in TopCase (McSherry, 2005), we find an example of a conversational system for tourism recommendations where users receive explanations of the recommendations if they ask.

Interactive. The explanation is already available to the user (in a proactive or reactive way), and the user can interact with the explanation to get more information. For example, Shopr (Lamche et al., 2014), allows users to interact with explanations to enhance their user experience.

5. Model Validation

The conceptual model has been created following an iterative design process, where every intermediate model was validated replicating the validation process used by Friedrich and Zanker (2011), He et al. (2016). The validation aims to determine if a validation set of explanation approaches can be classified with our model. If any of these approaches could not be classified, the model was refined. The process finished when all the approaches in the validation set were classified according to the proposed model.

The validation set comprises 50 explanation approaches and systems described in the literature. The criteria basis was to promote the heterogeneity of the explanation systems selected. We selected classic works and new approaches, traditional methods and innovative techniques. Although the validation set contains some explanation systems for decision-making systems, most of the approaches collected were explanations for recommender systems.

Next, we cite the collection of explanation systems employed in our validation in Table 2, and we discuss their properties according to our model.

The resulting classification is presented in Table 3, whose legend is shown in Table 4. Dash values ("-") represent that we do not have enough information to define its classification. Additionally, Figure 5 summarizes the number of instances categorized according to

System	Title & Reference
E_1	“Explainable Movie Recommendation Systems by using Story-based Similarity” (Lee & Jung, 2018)
E_2	“Explaining Recommendations by Means of User Reviews” (Donkers et al., 2018)
E_3	“Explaining Complex Scheduling Decisions” (Ludwig, Kalton, & Stottler, 2018)
E_4	“Explaining Contrasting Categories” (Pazzani, Feghahati, Shelton, & Seitz, 2018)
E_5	“Explaining smart heating systems to discourage fiddling with optimized behaviour” (Stumpf, Skrebe, Aymer, & Hobson, 2018)
E_6	“The design and validation of an intuitive confidence measure” (van der Waa, van Diggelen, & Neerincx, 2018)
E_7	“Interactive Explanations in Mobile Shopping Recommender Systems” (Lamche et al., 2014)
E_8, E_9, E_{10}	“A Review of Explanation and Explanation in Case-Based Reasoning” (Doyle et al., 2003)
E_{11}	“How should I explain? A comparison of different explanation types for recommender systems” (Gedikli et al., 2014)
$E_{12} - E_{18}$	“Explaining Collaborative Filtering Recommendations” (Herlocker et al., 2000).
E_{19}	“Tagsplanations: Explaining Recommendations Using Tags” (Vig et al., 2009)
E_{20}	“A case-based reasoning system for aiding detection and classification of nosocomial infections” (Gómez-Vallejo, Uriel-Latorre, Sande-Meijide, Villamarín-Bello, Pavón, Fdez-Riverola, & Glez-Peña, 2016)
E_{21}	“A framework for Explanation of Machine Learning Decisions” (Brinton, 2017)
E_{22}	“Great Explanations: Opinionated Explanations for Recommendations” (Muhammad et al., 2015)
E_{23}	“Knowledge-based systems, viewpoints and the world wide web” (Finch, 1998)
E_{24}	“More Than the Sum of Its Members: Challenges for Group Recommender Systems” (Jameson, 2004)
E_{25}	“Make it personal: A social explanation system applied to group recommendations” (Quijano-Sanchez et al., 2017)
E_{26}	“Crowd-based personalized natural language explanations for recommendations” (Chang, Harper, & Terveen, 2016)
E_{27}	“Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities” (He et al., 2016)
E_{28}	“Providing Control and Transparency in a Social Recommender System for Academic Conferences” (Tsai & Brusilovsky, 2017)
E_{29}	“Who to Follow and Why: Link Prediction with Explanations” (Barbieri et al., 2014)
E_{30}	“Explanation as a topic of AI research” (Buchanan & Shortliffe, 1984)
E_{31}	“Thinking Positively - Explanatory Feedback for Conversational Recommender Systems” (McCarthy, Reilly, McGinty, & Smyth, 2004)
E_{32}	“Trust-inspiring explanation interfaces for recommender systems” (Pu & Chen, 2007)
E_{33}	“Generating explanations in context: The system perspective” (Mittal & Paris, 1995)
E_{34}	“Do social Explanations Work? Studying and Modelling the effects of Social Explanations in Recommender Systems” (Sharma & Cosley, 2013)
E_{35}	“Providing Justifications in Recommender Systems” (Symeonidis, Nanopoulos, & Manolopoulos, 2008) and “MovieExplain: A recommender System with Explanations” (Symeonidis, Nanopoulos, & Manolopoulos, 2009)
E_{36}	“The Effectiveness of Personalized Movie Explanations: An Experiment Using Commercial Meta-data” (Tintarev & Masthoff, 2008)
E_{37}	“Explanation in recommender systems” (McSherry, 2005)
E_{38}	“Knowledgeable Explanations for Recommender systems” (Zanker & Ninaus, 2010)
E_{39}	“Explaining Recommendations Using Contexts” (Sato, Ahsan, Nagatani, Sonoda, Zhang, & Ohkuma, 2018)
E_{40}	“Exploring explanations for matrix factorization recommender systems” (Rastegarpanah, Crovella, & Gummadi, 2017)
E_{41}	“User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit” (Giboney, Brown, Lowry, & Nunamaker Jr, 2015)
E_{42}	“Tagcloud-based Explanation with Feedback for recommender systems” (Chen, Hsu, & Lee, 2013)
E_{43}	“ExpLOD: a Framework for Explaining Recommendations based on the Linked Open Data Cloud” (Musto et al., 2016)
E_{44}	“A Probabilistic Rating Prediction and Explanation Inference Model for Recommender Systems” (Wang et al., 2016)
E_{45}	“Providing Explanations for Recommendations in Reciprocal Environments” (Kleinerman, Rosenfeld, & Kraus, 2018)
E_{46}	“PRINCE: Provider-side Interpretability with Counterfactual Explanations in Recommender Systems” (Ghazimatin, Balalau, Saha Roy, & Weikum, 2020)
E_{47}	“A Model of Social Explanations for a Conversational Movie Recommendation System” (Pecune, Murali, Tsai, Matsuyama, & Cassell, 2019)
E_{48}	“Generating post hoc review-based natural language justifications for recommender systems” (Musto, de Gemmis, Lops, & Semeraro, 2020)
E_{49}	“Generate neural template explanations for recommendation” (Li, Zhang, & Chen, 2020)
E_{50}	“A Model-Agnostic Recommendation Explanation System Based on Knowledge Graph” (Chen & Miyazaki, 2020)

Table 2: Collection of explanation systems validated with the conceptual model.

SYS	Motivation			Knowledge		Generation		Presentation				
	USER	GOAL		DA	AL	TE	EN	FOR	ARG		DET	INT
		LL	TL						CP	PO		
E ₁	-	-	-	-	-	KI	PH	-	-	-	-	-
E ₂	-	T	-	BI	-	ML	PH	-	-	-	-	-
E ₃	TUS	-	-	C	-	KI	I	NL/V	SI	P	MTL	PR/RE
E ₄	TUS	-	-	C	-	ML	I	NL/V	SI	P	MTL	PR
E ₅	TUS	TR	-	PREF	-	KI	I	NL/SC/V	SI	P	MTL	PR/RE
E ₆	TUS	DEB	-	EX	-	KI	PH	SC	-	-	SL	-
E ₇	TUS	T/SCR	-	PREF	-	AH	PH	NL/V	SI/DI	P/N	MTL	PR/RE/IN
E ₈	TUS	-	-	EX	-	KI	I	NL	SI	P	MTL	PR
E ₉	TUS	-	-	EX	-	KI	I	SC	SI	P	-	PR
E ₁₀	-	-	-	EX	-	KI	I	-	-	-	-	-
E ₁₁	TUS	-	-	C/PREF	-	ML	PH	V	SI/DI	P/N	MTL	-
E ₁₂	TUS	EF/SA/T	-	-	US	ML	I/PH	V	SI	P	SL	-
E ₁₃	TUS	EF/SA/T	-	-	US	ML	I	V	SI	P	SL	-
E ₁₄	TUS	EF/SA/T	-	-	US	ML	I	SC	SI	P	SL	-
E ₁₅	TUS	EF/SA/T	-	-	-	ML	I	SC	SI	P	SL	-
E ₁₆	TUS	EF/SA/T	-	-	US	ML	I	SC	SI	P	SL	-
E ₁₇	TUS	EF/SA/T	-	COM	-	ML	I	SC	SI	P	SL	-
E ₁₈	TUS	EF/SA/T	-	COM	-	ML	I	SC	SI	P	SL	-
E ₁₉	TUS	T/EF/EFY	-	C/PREF	-	AH	PH	V	SI	P	MTL	-
E ₂₀	TUS	-	-	EX	-	KI	I	-	-	-	-	-
E ₂₁	-	-	-	C	-	AH	PH	NL	SI	P	SL	-
E ₂₂	TUS	EFY/PE	-	BI/EX/PREF	-	KI	I	SC	SI/DI	P/N	MTL	PR/IN
E ₂₃	TUS/STA	-	-	-	RP	KI	I	SC	SI/DI	P/N	SL	-
E ₂₄	TUG	-	-	PREF	-	KI	I	SC/V	SI	P	MTL	PR
E ₂₅	TUG	PE/EFY/TR	-	S/EX/BI	-	KI	PH	NL/V	SI/DI	P/N	MTL	-
E ₂₆	TUS/STA	EFY/EF/TR/SA	-	C/PREF	-	ML	PH	NL	SI	P	SL	-
E ₂₇	TUS	SA/TR/T	-	PREF/CO	US	-	-	SC/V	-	-	-	IN
E ₂₈	TUS	T	-	C	US/IS	AH	I	NL/SC/V	SI	P	MTL	IN
E ₂₉	-	-	-	C/S	-	AH	PH	-	SI	P	-	-
E ₃₀	TUS	-	-	-	RP	KI	I	NL	SI/DI	P/N	MTL	IN
E ₃₁	TUS	-	-	PREF	-	KI	PH	SC/V	SI/DI	P	MTL	PR/IN
E ₃₂	TUS	TR/EF/EFY	IU	C/PREF/COM	-	KI	I/PH	SC	SI/DI	P	MTL	IN
E ₃₃	TUS	EF	-	CO	RP	KI	I	NL	-	-	SL	PR/IN
E ₃₄	TUS	PE	-	S	-	ML	-	V	SI	P	SL	PR
E ₃₅	TUS	-	JR	PREF	-	ML	I	NL	SI	P	SL	-
E ₃₆	TUS	EF/TR/PE/T	-	COM/PREF/C	-	-	PH	NL	SI/DI	P/N	SL	-
E ₃₇	TUS	T	JR	PREF/EX	-	KI	I	NL	SI	P	SL	PR/IN
E ₃₈	TUS	T	-	PREF	-	KI	PH	NL	SI/DI	P	SL	-
E ₃₉	TUS	PE/EF	-	CO	-	MM	I	NL	SI	P	SL	-
E ₄₀	TUS	T	-	-	IS	MM	PH	SC	SI	P	SL	-
E ₄₁	TUS	EF	-	-	RP	KI	I	NL	SI	P	SL	PR
E ₄₂	TUS	EF/SA	-	C/PREF	-	ML	I	V	-	P/N	SL	PR
E ₄₃	TUS	T/PE/TR/SA/EF	-	PREF	IS	ML	PH	NL	SI	P	-	-
E ₄₄	TUS	EF/PE/SA	-	PREF/BI	IS	MM	I	-	SI	P	-	-
E ₄₅	TUG	SA/TR/T/EF	-	PREF	-	KI/MM	PH	NL/SC/V	SI	P	SL	PR
E ₄₆	TUS	SCR/EF	-	EX/C	-	KI	I	NL	SI	P	SL	-
E ₄₇	TUS	T/TR/EF/SA	IU	C	IS	KI	PH	NL	SI	P/N	MTL	RE/IN
E ₄₈	TUS	T/TR/EF	JR/IUE	C/CO	-	KI	PH	NL	SI	P	MTL	-
E ₄₉	TUS	-	JR	C	-	ML	I	NL	SI/DI	P/N	MTL	-
E ₅₀	TUS	EF	JR	C	IS	KI	PH	NL	SI	P	SL	-

Table 3: Classification of approaches studied according to our model

Concept	Term	Acronym	Concept	Term	Acronym
Motivation	System	SYS	Knowledge	Recommendation Process	RP
	Target User Single	TUS		Users Similarity	US
	Target User Group	TUG		Items Similarity	IS
	Stakeholders	STA	Generation	Technique	TE
	Developers	DEV		Encapsulation	EN
	Regulatory Body	BD		Introspective	I
	Top-Level	TL		Post hoc	PH
	Low-Level	LL		Knowledge-intensive	KI
	Transparency	T		Machine Learning	ML
	Scrutability	SCR		Mathematical Model	MM
	Trust	TR		Ad-hoc	AH
	Persuasiveness	PE		Format	FOR
	Effectiveness	EF		Natural Language	NL
	Efficiency	EFY	Schematic	SC	
	Satisfaction	SA	Visual	V	
	Education	ED	Other	O	
	Debugging	DEB	Argumentation	ARG	
	Improve User Retention	IU	Comparison	CP	
	Improve User Experience	IUE	Polarity	PO	
	Justify Recommendation	JR	Positive	P	
Knowledge	Algorithm	AL	Presentation	Negative	N
	Data	DA		Similarities	SI
	Content	C		Differences	DI
	Community	COM		Level of Detail	DET
	Preferences	PREF		Single-level	SL
	Background Information	BI		Multi-level	MTL
	Experiences	EX		Interaction	INT
	Context	CO		Proactive	PR
	Social	S		Reactive	RE
	Recommendation Process	RP		Interactive	IN

Table 4: Legend Table for Table 3

each concept defined in the ontology proposed. $E_5, E_7, E_{22}, E_{28}, E_{32}, E_{37}, E_{41}, E_{45}$ and E_{47} are the only approaches that can be classified in all the 4 explanation aspects defined by our model. We have to consider that systems E_{12} to E_{18} provide explanations for the same recommender; therefore, we will find a similar categorization.

Depending on the explanation’s motivation, in most of the approaches studied the user who receives the explanation is a single target user. We have found only two approaches whose user is a stakeholder (E_{23} and E_{26}), although they also explain decisions to a single target user. We found two examples of studies with a group as a type of user. In reference to the explanations’ main goal, we observed that many of the studies do not specify any. In addition, a few studies just describe their low-level goals. According to our results, the most usual goals are “Transparency” and “Effectiveness”.

Most of the studies inform about the knowledge used to generate the explanation. In the revisited publications, knowledge sources based on data are more used than algorithm-based ones. There are some obvious results in the table: when the system uses CBR, then the knowledge source is “Experiences” and, when the system uses collaborative filtering, then the most common knowledge source is “User similarity”. The “Content” type seems

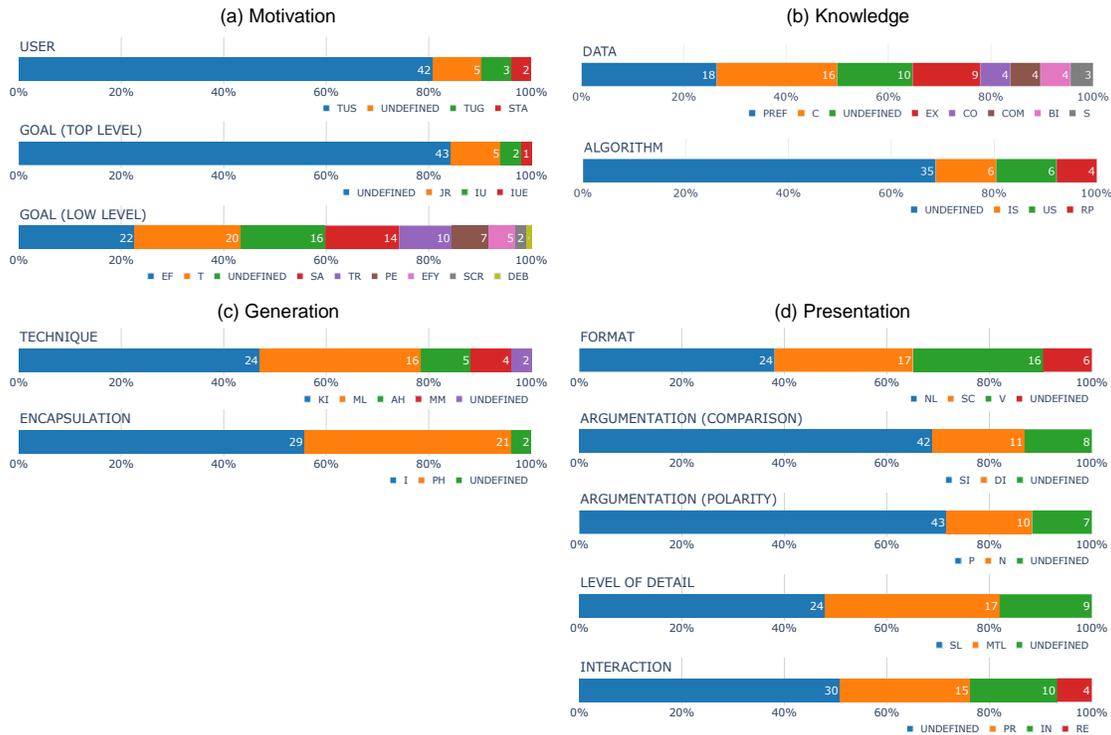


Figure 5: Number of instances classified according to the concepts and sub-concepts defined by our conceptual model

to be one of the most used knowledge sources, and it is commonly used in combination with “Preferences”. We found a few explanation approaches based on the remaining knowledge definition options, in an individual or combined way.

Two columns represent the “Generation” facet in Table 3 that summarize whether the system is introspective or post hoc (“Encapsulation” column) and the type of algorithm that generates the explanations (“Technique” column). We can see that the classification is equally distributed between the “Introspective” and the “Post hoc” values. Regarding the explanation algorithm, the most used is “Knowledge-intensive”, followed by “Machine Learning”. There are some systems whose algorithms are “Ad hoc”, and few systems using a “Mathematical model”.

In reference to the display format, we have found assorted results. Several studies present two formats: “Natural language” and “Visual”. In this case, natural language is used to explain by default, and visual, to explain with more details when users ask for it (reactive explanations). Regarding the argumentation facet, we can observe that in all of the methods studied, “Similarities” are commonly used with “Positive” argumentation, while “Differences” are mostly employed with “Negative” argumentation. We can also conclude that positive argumentation, with similarities, is the most widespread. On the other hand, the level of detail also varies. We have some very detailed explanation methods and other poorly detailed ones. Finally, we have not found too much information regarding

the interaction facet. We discovered that most of the “Proactive” methods are considered “Reactive” too. It should be noted that we have only found a few “Interactive” explanation approaches.

From the validation, we can state two main conclusions. On the one hand, our conceptual model fulfils the task of classifying the explanation systems and approaches in the validation set. On the other hand, the model contains a few facets that are not registered in Table 3. For example, we do not have any approach whose user is a regulatory body or a developer in Table 3. We do not have the education goal either. Although these facets are necessary in the design of an explanation system according to the more theoretical literature, we have not found any explanation approach that covers these facets during the literature review. Hence, we think that the proposed model is complete enough not only to classify the explanation systems encountered in the literature but to foresee the facets not covered yet by research.

6. Model Implementation

The goal of formalizing our conceptual model is to create an ontology employed during the development of explanation approaches for recommendation systems. Ontology-driven software development establishes the design and implementation of software using an ontology as guidance. The ontology itself works as a methodology to create the software (Pan, Staab, Aßmann, Ebert, & Zhao, 2012; Knublauch, 2004; Mavetera & Kroeze, 2010; Murtazina & Avdeenko, 2020). Our previous works in software composition applied to the CBR domain have shown the possibilities of the use of OWL-DL ontologies during the development process of CBR systems (Recio-García et al., 2014b; Díaz-Agudo, González-Calero, Recio-García, & Sánchez-Ruiz-Granados, 2007; Recio-García, Díaz-Agudo, & González-Calero, 2014a; Díaz-Agudo & González-Calero, 2002). These ontologies are used to describe systems and components, and their associated reasoning capabilities are exploited to aid the user in the generation of a complex system.

The model previously designed is formalized through the implementation of an ontology using the OWL language (McGuinness, Van Harmelen, et al., 2004), a Semantic Web language designed to represent rich and complex knowledge. We choose this language because of its popularity and its compatibility with the Description Logics (DL) reasoning of its OWL-DL sub-language. DLs are characterized by their expressiveness and clearly defined semantics, and they capture the meaning of the data by concentrating on entities (grouped into concepts) connected by relationships. More important than the DL’s representational characteristics are their reasoning mechanisms. These are based on *subsumption*, to infer whether one description is more general than another, and *instance recognition*, to infer the concepts that an individual satisfies and the relations that a tuple of individuals satisfies.

ExRecOnto ontology is the OWL-DL implementation of our conceptual model for designing explanations for recommender systems, presented in Section 4 and summarized in Figure 1. The interactive visualization of the ontology is available at <https://gaia.fdi.ucm.es/ontologies#exreconto>, and the source code is at: <https://github.com/martcaro/ExRecOnto>. In this ontology, we represent the hierarchy of concepts, the instances and the relationships among concepts and instances defined by our model. Next, we are going to present and explain these three components in our ontology.

Appendix Figures 6, 7, 8 and 9 show a graphical view of the four main concepts of the ontology: motivation, knowledge, generation and presentation. These figures also include the instances that belong to each concept. Leaving aside the standard relationships of any OWL ontology (*has-individual*, *has-subclass*, etc.) we defined the following relationships that are required to formalize our model. We show these relationships in Figure 10:

- *isComposedBy*. This considers the *Composition* relationship, which we have defined before in Section 3.
- *builds* and *isBuiltBy*. These are the relationships between the low-level and the top-level goals: the low-level goals of an explanation build top-level goals.
- *defines* and *dependsOn*. These are the relationships for User and Goal: the user defines the goal of the explanation.
- *delinates* and *showsAccordingTo*. The presentation of an explanation has a relationship with its goal: the goal defines the presentation.
- *determines* and *isDeterminedBy*. These define the relationships between the format and the knowledge of the explanation with the level of detail. The format and the knowledge determine the level of detail.
- *influences* and *isInfluencedBy*. The goal of the explanation influences how the explanation is argued.
- *isBasedOn* and *isBaseOf*. These relationships show the connection between our facet generation and algorithm: the explanation algorithm is based on how the explanation is generated.
- *isUsedBy* and *uses*. The implementation of the explanation uses the data available for its generation.

7. Conclusions

This work presents a thorough revision of state of the art on explanations for recommender systems. From this analysis, a novel conceptual model is designed to integrate and enhance existing approaches. This model analyzes and categorizes explanations from diverse points of view, considering features and concepts of existing works but enhanced with additional aspects and the formalization knowledge that other models lack.

Our model proposes and details four main aspects that must be taken into account when analyzing or generating explanations for recommendation systems: the user's motivation and goals, the knowledge required by its generation, the recommendation process itself and the presentation to the user. In this work, we detail these aspects and define their dependencies. We have collected an extensive list of explanation approaches from the literature and classified them according to our model to validate the proposal.

However, the model might have some limitations. XAI (eXplainable Artificial Intelligence) is one of the most on-topic research fields in computer science nowadays, and it is constantly changing. Therefore, it is to be expected that pioneer approaches might appear.

Although the conceptual model foresees some interesting facets not covered by previous explanation approaches, maybe new approaches involve new innovative aspects not covered by the current version of the conceptual model. However, the iterative design process followed during the creation of the conceptual model can be applied again to enhance it with the new facets that might be yet to come.

Nevertheless, we expect this model to be a useful tool to guide the design and development of explanations for recommender systems in most cases. It provides a broader picture of the diversity of facets that must be considered during an explanation system's development process. To do so, we provide a formalization of the conceptual model as an ontology *ExRecOnto*, using the OWL semantic web language, which is publicly available. According to ontology-driven software development, this ontology can guide the integration of explanations into existing frameworks to develop recommender systems.

Appendix A: Ontology Visualization

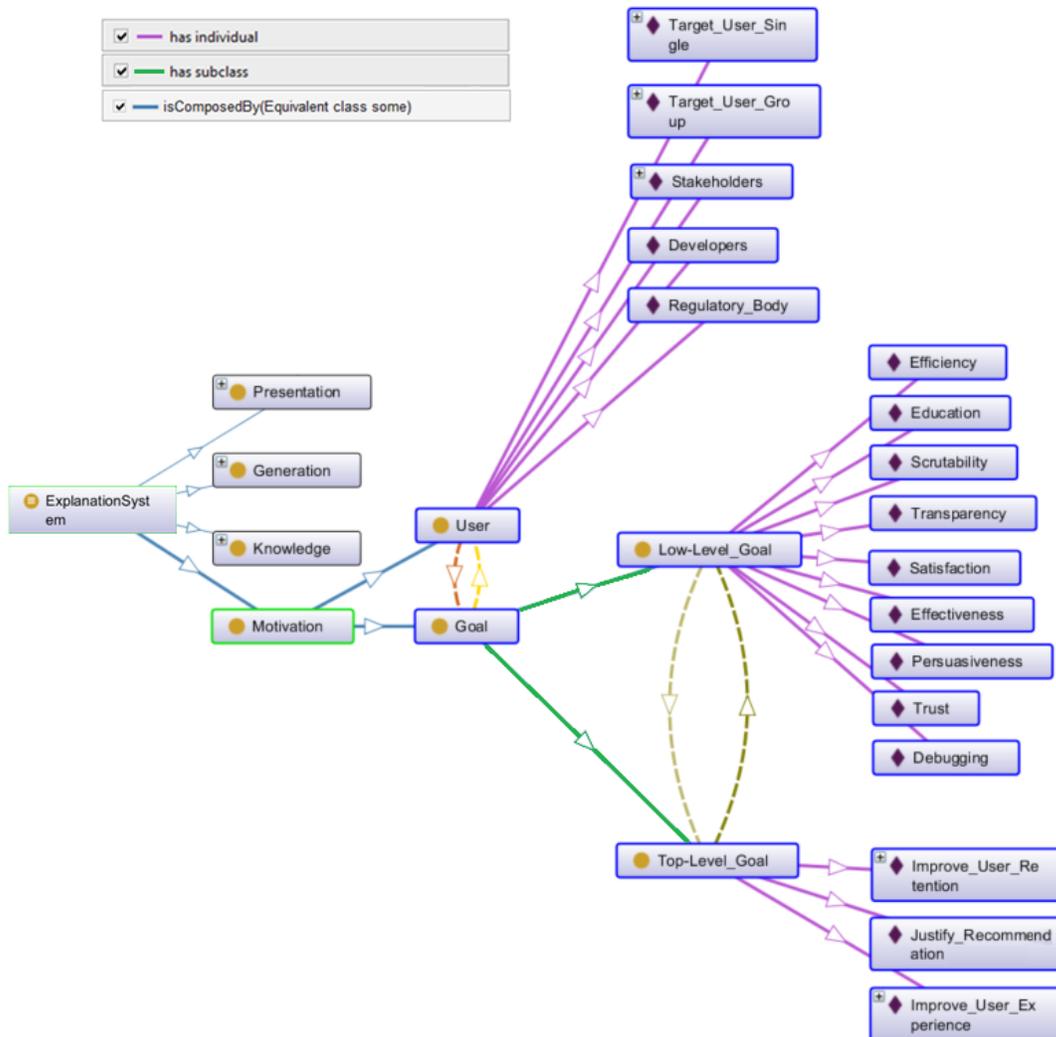


Figure 6: Partial view of the ontology corresponding to the *motivation* concept

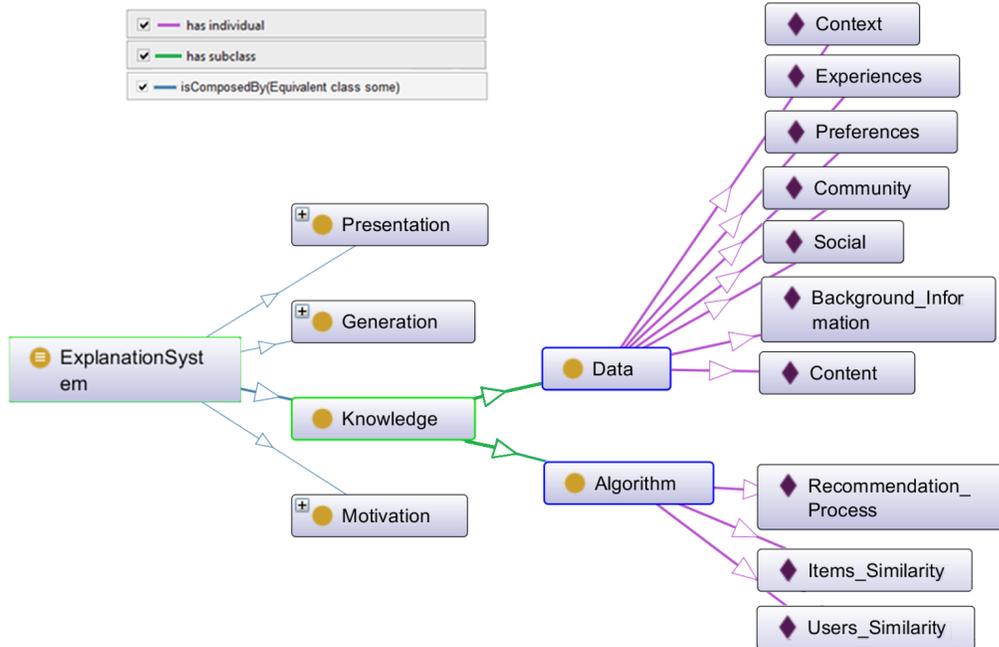


Figure 7: Partial view of the ontology corresponding to the *knowledge* concept

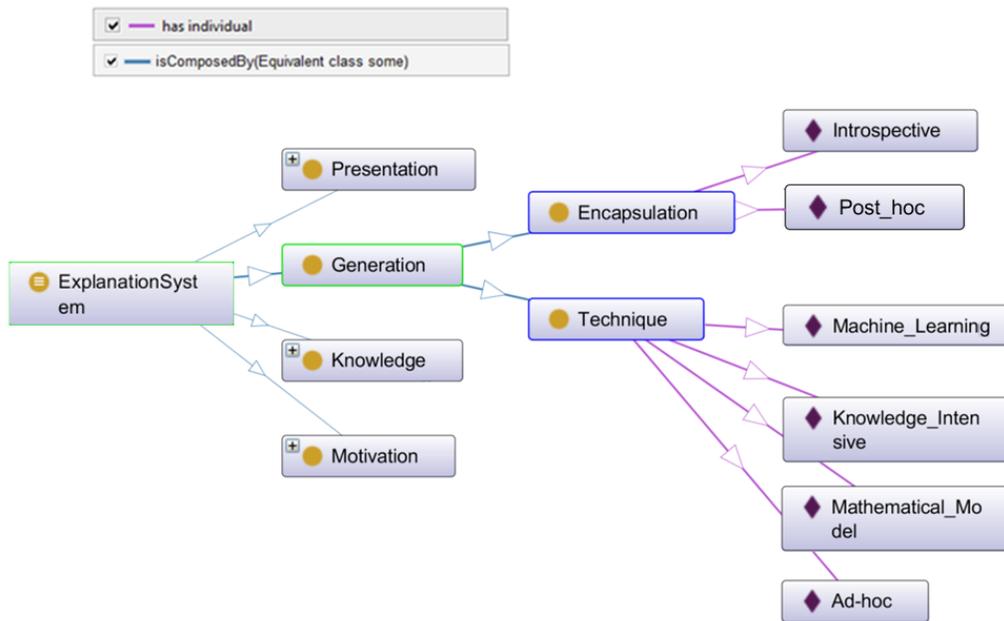


Figure 8: Partial view of the ontology corresponding to the *generation* concept

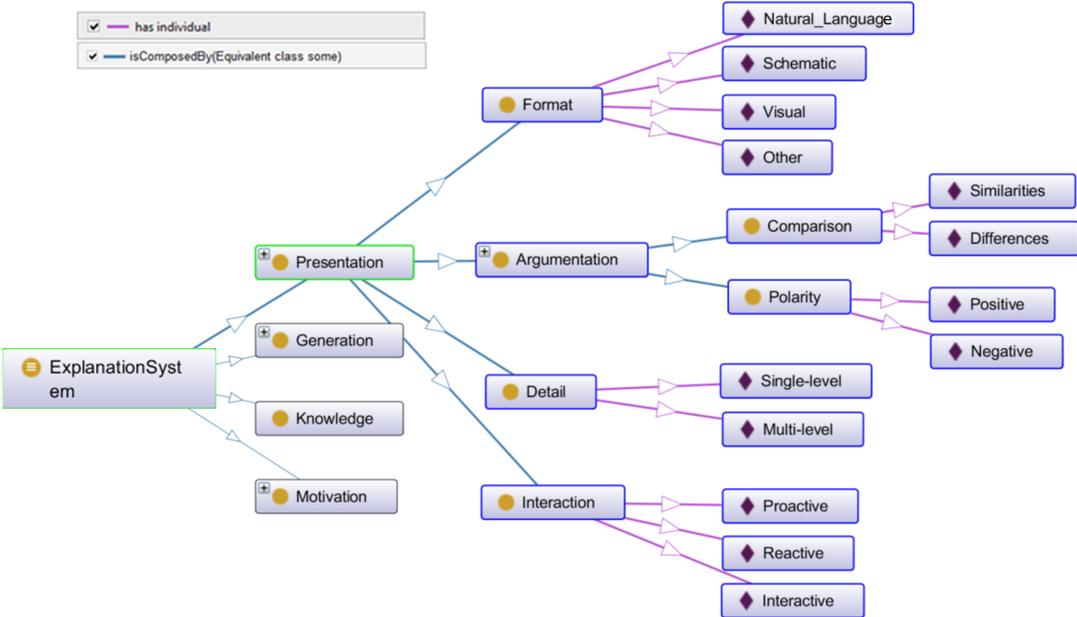


Figure 9: Partial view of the ontology corresponding to the *presentation* concept

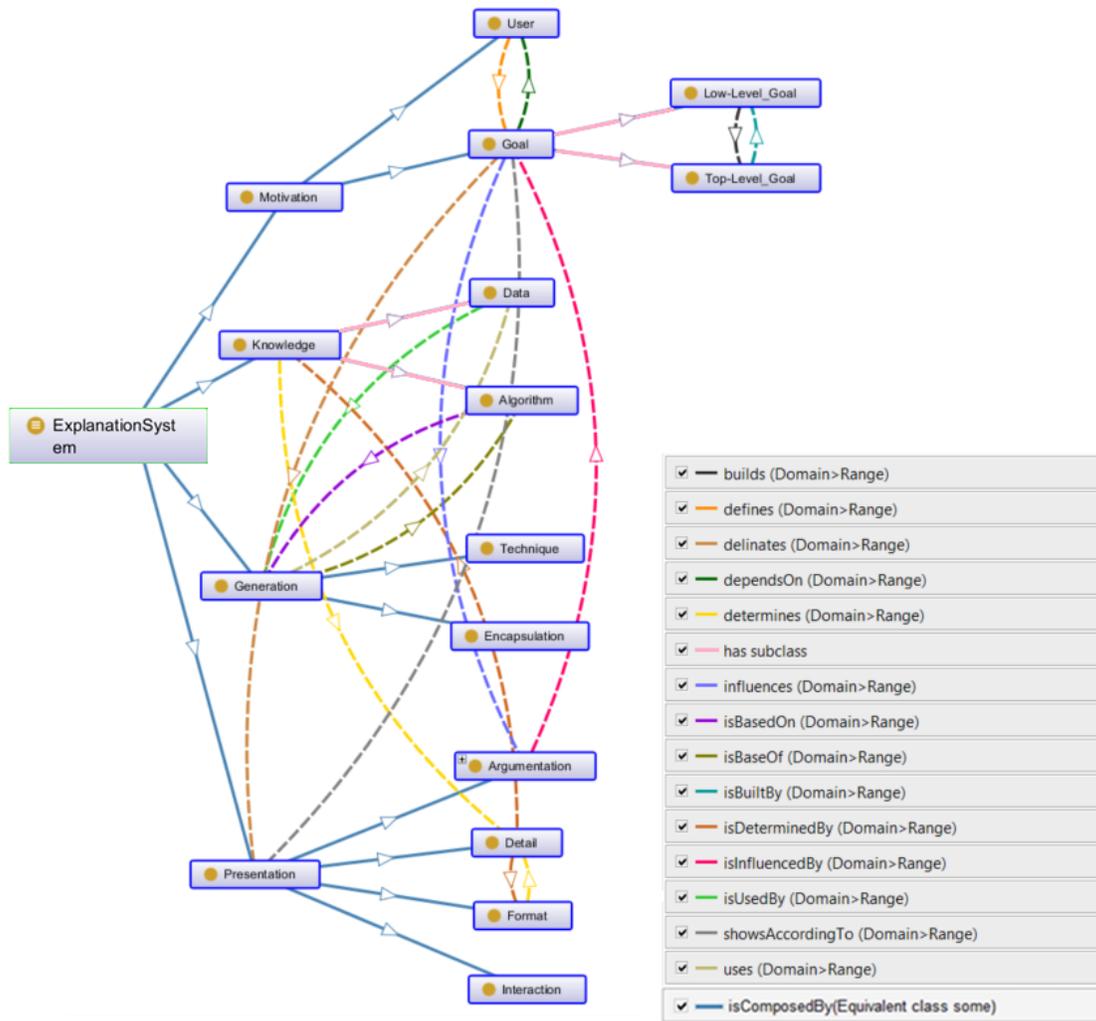


Figure 10: View of the relationships between the ontology concepts

References

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6, 52138–52160.
- Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook*, pp. 217–253. Springer.
- Al-Taie, M. Z., & Kadry, S. (2014). Visualization of explanations in recommender systems. *Journal of Advanced Management Science Vol*, 2(2).
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks...
- Arrieta, A. B., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58, 82–115.
- Barbieri, N., Bonchi, F., & Manco, G. (2014). Who to follow and why: link prediction with explanations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1266–1275. ACM.
- Berkovsky, S., Taib, R., & Conway, D. (2017). How to recommend?: User trust factors in movie recommender systems. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 287–300. ACM.
- Brinton, C. (2017). A Framework for explanation of machine learning decisions. In *IJCAI-17 Workshop on Explainable AI (XAI)*, p. 14.
- Buchanan, B., & Shortliffe, E. (1984). The MYCIN experiments of the stanford heuristic programming project. *Reading, MA: Addison-Wasley*.
- Buchanan, B. (1984). Rule based expert systems. *The MYCIN Experiments of the Stanford Heuristic Programming Project*.
- Caro-Martinez, M., Jimenez-Diaz, G., & Recio-Garcia, J. A. (2018). A theoretical model of explanations in recommender systems. *ICCBR 2018*, 52.
- Chang, S., Harper, F. M., & Terveen, L. G. (2016). Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 175–182. ACM.
- Chen, W., Hsu, W., & Lee, M. L. (2013). Tagcloud-based explanation with feedback for recommender systems. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 945–948. ACM.
- Chen, Y., & Miyazaki, J. (2020). A model-agnostic recommendation explanation system based on knowledge graph. In *International Conference on Database and Expert Systems Applications*, pp. 149–163. Springer.
- Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.
- Díaz-Agudo, B., González-Calero, P. A., Recio-García, J. A., & Sánchez-Ruiz-Granados, A. A. (2007). Building CBR systems with jCOLIBRI. *Science of Computer Programming*, 69(1-3), 68–75.

- Díaz-Agudo, B., & González-Calero, P. A. (2002). CBROnto: a task/method ontology for CBR. *Procs. of the 15th International FLAIRS, 2*, 101–106.
- Donkers, T., Loepp, B., & Ziegler, J. (2018). Explaining Recommendations by Means of User Reviews. In *Workshop On Explainable Smart Systems (EXSS)*.
- Doyle, D., Tsymbal, A., & Cunningham, P. (2003). A review of explanation and explanation in case-based reasoning. Tech. rep., Trinity College Dublin, Department of Computer Science.
- Eiband, M., Schneider, H., & Buschek, D. (2018). Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems. In *Workshop On Explainable Smart Systems (EXSS)*.
- Euzenat, J., Shvaiko, P., et al. (2007). *Ontology matching*, Vol. 18. Springer.
- Finch, I. (1998). Knowledge-based systems, viewpoints and the world wide web.. IET.
- Friedman, A., Knijnenburg, B. P., Vanhecke, K., Martens, L., & Berkovsky, S. (2015). Privacy aspects of recommender systems. In *Recommender Systems Handbook*, pp. 649–688. Springer.
- Friedrich, G., & Zanker, M. (2011). A taxonomy for generating explanations in recommender systems. *AI Magazine, 32*(3), 90–98.
- Gavalas, D., Konstantopoulos, C., Mastakas, K., & Pantziou, G. (2014). Mobile recommender systems in tourism. *Journal of network and computer applications, 39*, 319–333.
- Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies, 72*(4), 367–382.
- Ghazimatin, A., Balalau, O., Saha Roy, R., & Weikum, G. (2020). Prince: Provider-side interpretability with counterfactual explanations in recommender systems. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 196–204.
- Giboney, J. S., Brown, S. A., Lowry, P. B., & Nunamaker Jr, J. F. (2015). User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit. *Decision Support Systems, 72*, 1–10.
- Gómez-Vallejo, H., Uriel-Latorre, B., Sande-Meijide, M., Villamarín-Bello, B., Pavón, R., Fdez-Riverola, F., & Glez-Peña, D. (2016). A case-based reasoning system for aiding detection and classification of nosocomial infections. *Decision Support Systems, 84*, 104–116.
- Green, H. E. (2014). Use of theoretical and conceptual frameworks in qualitative research. *Nurse researcher, 21*(6).
- He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications, 56*, 9–27.

- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp. 241–250. ACM.
- Jameson, A. (2004). More than the sum of its members: challenges for group recommender systems. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 48–54. ACM.
- Jameson, A., & Smyth, B. (2007). Recommendation to groups. In *The adaptive web*, pp. 596–627. Springer.
- Jorro-Aragoneses, J. L., Ceron-Rios, G. M., Díaz-Agudo, M. B., Recio-García, J. A., & López-Gutierrez, D. M. (2017). RecOnto: An ontology to model recommender systems and its components. In *Tools with Artificial Intelligence (ICTAI), 2017 IEEE 29th International Conference on*, pp. 815–821. IEEE.
- Kleinerman, A., Rosenfeld, A., & Kraus, S. (2018). Providing explanations for recommendations in reciprocal environments. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 22–30. ACM.
- Knublauch, H. (2004). Ontology-driven software development in the context of the semantic web: An example scenario with protege/owl. In *1st International workshop on the model-driven semantic web (MDSW2004)*, pp. 381–401. Monterey, California, USA.
- Lamche, B., Adıgüzel, U., & Wörndl, W. (2014). Interactive explanations in mobile shopping recommender systems. In *Joint Workshop on Interfaces and Human Decision Making in Recommender Systems*, p. 14.
- Lee, O.-J., & Jung, J. J. (2018). Explainable Movie Recommendation Systems by using Story-based Similarity. In *Workshop On Explainable Smart Systems (EXSS)*.
- Li, L., Zhang, Y., & Chen, L. (2020). Generate neural template explanations for recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 755–764.
- Lu, Y., Dong, R., & Smyth, B. (2018). Coevolutionary recommendation model: Mutual learning between ratings and reviews. In *Proceedings of the 2018 World Wide Web Conference*, pp. 773–782.
- Ludwig, J., Kalton, A., & Stottler, R. (2018). Explaining Complex Scheduling Decisions. In *Workshop On Explainable Smart Systems (EXSS)*.
- Mavetera, N., & Kroeze, J. H. (2010). An ontology-driven software development framework..
- McCarthy, K., Reilly, J., McGinty, L., & Smyth, B. (2004). Thinking positively-explanatory feedback for conversational recommender systems. In *Proceedings of the European Conference on Case-Based Reasoning (ECCBR-04) Explanation Workshop*, pp. 115–124.
- McGuinness, D. L., Van Harmelen, F., et al. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10), 2004.
- McSherry, D. (2005). Explanation in recommender systems. *Artificial Intelligence Review*, 24(2), 179–197.

- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.
- Mittal, V. O., & Paris, C. L. (1995). Generating explanations in context: The system perspective. *Expert Systems with Applications*, 8(4), 491–503.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.
- Muhammad, K., Lawlor, A., Rafter, R., & Smyth, B. (2015). Great explanations: Opinionated explanations for recommendations. In *International Conference on Case-Based Reasoning*, pp. 244–258. Springer.
- Murtazina, M., & Avdeenko, T. (2020). The ontology-driven approach to intelligent support of requirements engineering in agile software development. In *2020 International Conference on Information Technology and Nanotechnology (ITNT)*, pp. 1–6. IEEE.
- Musto, C., de Gemmis, M., Lops, P., & Semeraro, G. (2020). Generating post hoc review-based natural language justifications for recommender systems. *User Modeling and User-Adapted Interaction*, 1–45.
- Musto, C., Narducci, F., Lops, P., De Gemmis, M., & Semeraro, G. (2016). ExpLOD: a framework for explaining recommendations based on the linked open data cloud. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 151–154. ACM.
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5), 393–444.
- Pan, J. Z., Staab, S., Abmann, U., Ebert, J., & Zhao, Y. (2012). *Ontology-driven software development*. Springer Science & Business Media.
- Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2012). A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, 24(3), 555–583.
- Parra, D., Brusilovsky, P., & Trattner, C. (2014). See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 235–240.
- Pazzani, M., Feghahati, A., Shelton, C., & Seitz, A. (2018). Explaining Contrasting Categories. In *Workshop On Explainable Smart Systems (EXSS)*.
- Pecune, F., Murali, S., Tsai, V., Matsuyama, Y., & Cassell, J. (2019). A model of social explanations for a conversational movie recommendation system. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pp. 135–143.
- Peña, F. J., O'Reilly-Morgan, D., Tragos, E. Z., Hurley, N., Duriakova, E., Smyth, B., & Lawlor, A. (2020). Combining rating and review data by initializing latent factor models with topic models for top-n recommendation. In *Fourteenth ACM Conference on Recommender Systems*, pp. 438–443.
- Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6), 542–556.

- Quijano-Sanchez, L., Sauer, C., Recio-Garcia, J. A., & Diaz-Agudo, B. (2017). Make it personal: a social explanation system applied to group recommendations. *Expert Systems with Applications*, *76*, 36–48.
- Rastegarpanah, B., Crovella, M., & Gummadi, K. P. (2017). Exploring Explanations for Matrix Factorization Recommender Systems..
- Recio-García, J. A., Díaz-Agudo, B., & González-Calero, P. A. (2014a). The COLIBRI platform: tools, features and working examples. In *Successful Case-based Reasoning Applications-2*, pp. 55–85. Springer.
- Recio-García, J. A., González-Calero, P. A., & Díaz-Agudo, B. (2014b). Template-based design in colibri studio. *Information Systems*, *40*, 168–178.
- Recio-Garcia, J. A., Jimenez-Diaz, G., Sanchez-Ruiz, A. A., & Diaz-Agudo, B. (2009). Personality aware recommendations to groups. In *Proceedings of the third ACM conference on Recommender systems*, pp. 325–328. ACM.
- Richter, M. M., & Weber, R. O. (2016). *Case-based reasoning*. Springer.
- Robillard, M., Walker, R., & Zimmermann, T. (2010). Recommendation systems for software engineering. *IEEE software*, *27*(4), 80–86.
- Ruiz-Iniesta, A., Jimenez-Diaz, G., & Gomez-Albarran, M. (2009). Recommendation in repositories of learning objects: A proactive approach that exploits diversity and navigation-by-proposing. In *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on*, pp. 543–545. IEEE.
- Sansonetti, G., Gasparetti, F., Micarelli, A., Cena, F., & Gena, C. (2019). Enhancing cultural recommendations through social and linked open data. *User Modeling and User-Adapted Interaction*, *29*(1), 121–159.
- Sato, M., Ahsan, B., Nagatani, K., Sonoda, T., Zhang, Q., & Ohkuma, T. (2018). Explaining Recommendations Using Contexts. In *23rd International Conference on Intelligent User Interfaces*, pp. 659–664. ACM.
- Sharma, A., & Cosley, D. (2013). Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1133–1144. ACM.
- Stumpf, S., Skrebe, S., Aymer, G., & Hobson, J. (2018). Explaining smart heating systems to discourage fiddling with optimized behavior. In *Workshop On Explainable Smart Systems (EXSS)*.
- Symeonidis, P., Nanopoulos, A., & Manolopoulos, Y. (2008). Providing justifications in recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *38*(6), 1262–1272.
- Symeonidis, P., Nanopoulos, A., & Manolopoulos, Y. (2009). MoviExplain: a recommender system with explanations. In *Proceedings of the third ACM conference on Recommender systems*, pp. 317–320. ACM.
- Terano, T., Suzuki, M., Onoda, T., Uenishi, K., & Matsuura, T. (1989). CSES: an approach to integrating graphic, music and voice information into a user-friendly interface.

- In *Industrial Applications of Machine Intelligence and Vision, 1989., International Workshop on*, pp. 349–354. IEEE.
- Tintarev, N., & Masthoff, J. (2008). The effectiveness of personalized movie explanations: An experiment using commercial meta-data. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 204–213. Springer.
- Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 399–439.
- Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*, pp. 353–382. Springer.
- Tsai, C.-H., & Brusilovsky, P. (2017). Providing Control and Transparency in a Social Recommender System for Academic Conferences. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 313–317. ACM.
- Tsai, C.-H., & Brusilovsky, P. (2020). The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction*, 1–37.
- van der Waa, J., van Diggelen, J., & Neerincx, M. (2018). The design and validation of an intuitive confidence measure. In *Workshop On Explainable Smart Systems (EXSS)*, Vol. 2, p. 1.
- Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (2013). Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pp. 351–362.
- Vig, J., Sen, S., & Riedl, J. (2009). Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pp. 47–56. ACM.
- Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31, 841.
- Wang, H., Fi, Q., Liu, L., & Song, W. (2016). A probabilistic rating prediction and explanation inference model for recommender systems. *China Communications*, 13(2), 79–94.
- Zanker, M., & Ninaus, D. (2010). Knowledgeable explanations for recommender systems. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, Vol. 1, pp. 657–660. IEEE.
- Zhang, Y., & Chen, X. (2018). Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*.