A Holistic Approach for Semantic-Based Game Generation

Owen Sacco
Institute of Digital Games
University of Malta
Msida, Malta
owensacco@gmail.com

Antonios Liapis
Institute of Digital Games
University of Malta
Msida, Malta
antonios.liapis@um.edu.mt

Georgios N. Yannakakis
Institute of Digital Games
University of Malta
Msida, Malta
georgios.yannakakis@um.edu.mt

Abstract—The Web contains vast sources of content that could be reused to reduce the development time and effort to create games. However, most Web content is unstructured and lacks meaning for machines to be able to process and infer new knowledge. The *Web of Data* is a term used to describe a trend for publishing and interlinking previously disconnected datasets on the Web in order to make them more valuable and useful as a whole. In this paper, we describe an innovative approach that exploits Semantic Web technologies to automatically generate games by reusing Web content. Existing work on automatic game content generation through algorithmic means focuses primarily on a set of parameters within constrained game design spaces such as terrains or game levels, but does not harness the potential of already existing content on the Web for game generation. We instead propose a holistic and more generally-applicable game generation solution that would identify suitable Web information sources and enrich game content with semantic meta-structures.

I. INTRODUCTION

Creating deep and largely non-linear games that are market-competent costs in time, effort and resources which often cannot be afforded by small-medium enterprises, especially by independent game development studios. As most of the tasks involved in developing games are labour- and creativity-intensive, our vision is to reduce software development effort and enhance design creativity by automatically generating novel and semantically-enriched content for games from Web sources. In this paper we envision a game generator that extracts information directly from the Web such as from wiki articles or images that are freely available. Following a semantic-based game generation approach not only can reduce the time and cost of game content creation but also directly contribute to web-informed yet unconventional game design.

Suppose a system where a user enters a description of a video game that s/he desires. The description could contain some keywords such as "*action adventure video game*", or could contain a complete narrative that describes the whole plot of the game. The system would then generate a video game based on the user's input by reusing Web content. Advanced features would be provided such as properties that the user can select prior to the system generating the game. An editor would also be provided for modifying the video game prototype after it is generated. All content generated from this system would be semantically enriched described

using standard structured meta-formats that would enable the content to be re-publishable on the Web for easy interlinking and consumption. This system would be beneficial for non-technical users without any background in developing digital games. For example by using this system, educators can generate educational games for their students about a particular subject without having to develop the game. Another example, health practitioners can generate a game by using this system for their patients to treat a particular disease. Moreover, this system would also be beneficial for experienced developers since it would reduce the time and effort for them to develop games by generating an unconventional set of playable game prototypes. Furthermore, our approach would publish semantically enriched game information that can be reused in other games.

Games are composed of different domains (or *facets*) that contribute to the game's look, feel and experience [24]. These facets include visuals, audio, narrative, gameplay, game design and level design. Each facet can be regarded as an independent model containing specific content, and a game is created when each of these models are interlinked together based on the game's requirements. Current work on automatic generation of content comprise of algorithms that generate limited in-game entities, such as *SpeedTree* [39] that generates trees and vegetation as part of the visuals facet, or the Ludi system [7] which generates game rules for two-player board games as part of the game design facet. Although such algorithms are beneficial for automatic generation of content, it is still rare that more than one domain is considered — e.g. in the work of Cook [10], Lopes et al. [26] and Riedl et al. [18] — requiring a substantial amount of manual effort by game developers to create content.

Inclusion of digital real-world data (e.g. recent news, real landscapes or historical events) in game environment is a practice used to increase the reality of scenarios and game play. Many flight simulators (e.g. *Flight Simulator X* (Microsoft Game Studios, 2006) or *Flight Pro Sim*[1]) include digital models of objects such as actual landscapes or airports and combine them with multiplayer mode to provide more realistic experiences. This trend is also visible in some of the sports

---

[1]Flight Pro Sim — http://flightprosim.com

games, such as the *SSX* series of games (EA Sports, 2000) where the NASA topography was used for the creation of snowboarding trails [30]. The inclusion of historical events is especially visible in strategic games where players can participate in scenarios resembling actual campaigns and immerse in history such as the *Call of Duty* (Activision, 2003) or the *Medal of Honor* (Electronic Arts, 1999) series. Moreover, games often include historic information within the plot that gamers can interact with, such as in the *Assassin's Creed* (Ubisoft, 2007) series, which contributes to a more immersive gaming experience. However, most of this content is currently created manually even though such information is widely available on the Web.

Web content is dispersed over the Internet in the form of blogs, microblogs, forums, wikis, social networks, review sites, and other Web applications which are currently disconnected from one another. The datasets created by these communities all contain information which can be used to generate or reuse content in games, but are not easily discoverable. The emerging Web of Data trend [6], where datasets are published in a standard form for easy interlinking, enables to essentially view the whole Web as one massive integrated database. Nevertheless, game information is still not enriched with meta-structures that could be used both on the Web and also in games. With such rich meta-structures that add more meaning to content, this would enable Web content to be reused in games. Moreover, the representation of *semantically-enriched* and *semantically-interlinked* content would enable game generators to infer how content can be interacted within the game world without having to rely on software development procedures that require laborious annotation of how each entity can be interacted within the game.

In this paper, we introduce an approach that automatically generates games from Web content through the use of rich-meta structures to describe game content. This research direction would provide a standard format for structuring and describing content for each game facet that would in turn be interlinked to automatically generate games. Our approach uses content from review gaming sites, game ranking sites and walkthroughs to generate games. This enriched content would be represented as interlinked game facet graphs containing information including visuals, audio, gameplay, rules, levels, user profiles, character information, and other relevant game information that could be extracted from diverse Web sources. The content is extracted and semantically enriched with meta-structures using new and/or already existing Semantic Web ontologies. Among other applications, our model could be used: (1) to define meta-structures for characterising and representing game data abstractly that could then be re-used on the Web from games; and (2) to integrate game content from the Web within games. This innovative way of game creation could lead to a new kind of real-time game experience which also implies the design of new game play and rules based on semantic information. The game generation approach envisaged is ambitious but nevertheless feasible with current technology as discussed in the remaining sections of this paper.

The remainder of this paper is as follows. Section II reviews current work on procedural game generation and semantics in games and Section III offers core background information about the Web of data. In Section IV we detail our approach of generating games using semantic data and Section V concludes the paper by providing an overall discussion about the future steps of this work.

II. Procedural Content Generation and Semantics

Development and research on automatic content generation mainly focuses on procedural content generation (PCG) that refers to the creation of game content automatically via algorithmic means [35]. Game content, primarily levels and visual assets, have been generated algorithmically in the game industry for decades. From the dungeons of *Rogue* (Toy and Wichman, 1980) to the universe of *Elite* (Acornsoft, 1984), procedural content generator has largely focused on the spatial structures of a specific game (with a specific, human authored ruleset and narrative). However, recent advances in PCG in academia has targeted a broader range of game facets: from the game rules [7] to soundscapes [27], and from the story [32] to the game's shaders [20], there is significant potential in computational creativity [24] in game design that has not been considered previously. Even though some generators have used real-world information as a starting point (seed) or a post-processing step (decoration) of their generative process most types of game content are not generated based on real-world information — or more appropriately, the human engineers insert their real-world assumptions (e.g. on what constitutes a "valid" castle) into the generator.

Angelina [11] uses *Guardian* articles as a seed for a platformer level: the textual information is parsed (extracting nouns) and using sentiment word analysis the "tone" of the article is computed (positive or negative). The platformer level is then decorated (i.e. colors, backgrounds, sound effects) based on the topics and tone, via Google search. *Game-O-Matic* [36] on the other hand relies on human-provided entities and their interactions (ways in which an entity affects another, such as "Hunger harms man") to generate simple games. On the other hand, the games' mechanics are based on combined game behaviours which are provided by the generator's designers to simulate the interactions. The in-game entities' appearance are found via Google image search, and rendered as such in-game. It is obvious that in these examples, the real-world information provided by the Web mostly act as decoration to generated content (platformer levels or arcade games) which are created without influence from real-world data. Instead, the proposed approach goes beyond merely decorating known good generated results but instead integrates the real-world data more in all facets of generation. Another approach is *A Rogue Dream* [9], which uses a single word from the user (acting as the identity of the player's avatar, such as cat) and uses the auto-complete function of Google queries (e.g. "why do cats hate...") to find the semantic identity of enemies, goals and special abilities of the player avatar.

The semantic identity (e.g. "dogs" in the above example) is depicted visually via Google image search. A Rogue Dream uses the associations existent in the communal knowledge pool (as popular search topics) but does not rely on structured data, as the current paper proposes; instead, it directly parses the text of autocompleted Google queries without, for instance, checking whether the potential enemy is indeed of type "living creature" or "person".

Perhaps the closest attempt at using structured real-world data is the Data Adventures project [2], [3] which uses SPARQL queries (see Section III for a detailed description of SPARQL) on DBpedia to discover links between two or more individuals: the discovered links are transformed into adventure games, with entities of type "Person" becoming Non-Player Characters (NPCs) that the player can converse with, entities of type "City" becoming cities that the player can visit, and entities of type "Category" becoming books that the player can read. The proposed approach enhances the concepts explored in Data Adventures by using game information (from real-world games such as those found in Metacritic) in order to discover possible new mechanics: Data Adventures instead relies on hand-authored mechanics existent in traditional adventure games of the 1980s.

Most of these methods do not reuse already generated content from the Web and do not add semantics to the generated content which could be used by the game engine to infer (new) knowledge. Games could take any form in real-time based on semantic information, and semantic structures for game information create standard ways for publishing and reusing content in many games.

The advantages of using rich semantic information to automatically generate games are numerous [37] as more complex, open-world, non-linear, games incorporating very rich forms of interaction are possible (i.e. authentic sandbox games). Current work in using semantics in games focuses on the use of semantic information to generate game worlds or to describe interactions with game worlds such as the work in [21], [28], [38]. Although these provide useful insights in generic semantic models that describe interactions with game worlds, they do not offer vocabularies for describing game content and they neither provide a generic approach for reusing Web content to generate games.

Attempts in game ontology creation are relevant to our approach, hence, we outline the four key game-based ontologies existent currently. The Game Ontology Project [43] is a wiki-based knowledge-base that aims to provide elements of gameplay. However, this project does not take into consideration game content and does not provide fine-grained concepts that cover different aspects of information within the different game facets. Moreover, it does not provide a vocabulary to be consumed by data described in RDF which could make it potentially useful for game generation. The Digital Game Ontology [8] provides an ontology by aligning with the Music Ontology, and the Event and Timeline ontology, to provide concepts that describe digital games. However, the vocabulary is not available and in this regard, it is unclear what game concepts this vocabulary provides. Finally, the Video Game Ontology [31] provides concepts for defining interoperability amongst video games and the Game2Web ontology [33] focuses on linking game events and entities to social data. Although these vocabularies are useful for describing several aspects of game information the ontologies are still limited to specific features of particular facets of game generation.

III. BACKGROUND: THE WEB OF DATA

The *Web of Data* is evolving the Web to be consumed both by machines and humans whereas the traditional Web resulted to be for human consumption only. Indeed, machines cannot process additional meaning from the content found in Web pages since they are simply text and similarly from the non-typed links which do not contain any additional meaning about the relationships amongst the linked pages. Therefore, the *Web of Data* provides various open data formats which have emerged from the Semantic Web.

A. The Semantic Web

The Semantic Web [5] provides approaches for structuring information on the Web by using metadata to describe Web data. The advantage of using metadata is that information is added with meaning whereby Web agents or Web enabled devices can process such meaning to carryout complex tasks automatically on behalf of users. Another advantage is that the semantics in metadata improved the way information is presented, for instance merging information from heterogeneous sources on the basis of the relationships amongst data, even if the underlying data schemata differ. Therefore, the Semantic Web encouraged the creation of meta-formats to describe metadata that can be processed by machines to infer additional information, to allow for data sharing and to allow for interoperability amongst Web pages. The common format and recommended by W3C for Semantic data representation [4] is the Resource Description Framework (RDF)[2].

B. Resource Description Framework (RDF)

RDF is a framework that describes resources on the World Wide Web. Resources can be anything that can be described on the Web; being real-world entities such as a person, real-world objects such as a car and abstract concepts such as defining the concept of game review scores. RDF provides a framework for representing data that can be exchanged without loss of meaning. RDF uniquely identifies resources on the Web by means of Uniform Resource Identifiers (URIs). Resources are described in RDF in the form of triple statements. A triple statement consists of a *subject*, a *predicate* and an *object*. A subject consists of the unique identifier that identifies the resource. A predicate represents the property characteristics of the subject that the resource specifies. An object consists of the property value of that statement. Values can be either literals or other resources. Therefore, the predicate of the RDF statement describes relationships between the subject and the object. If a triple had to be depicted as a graph, the subject

---

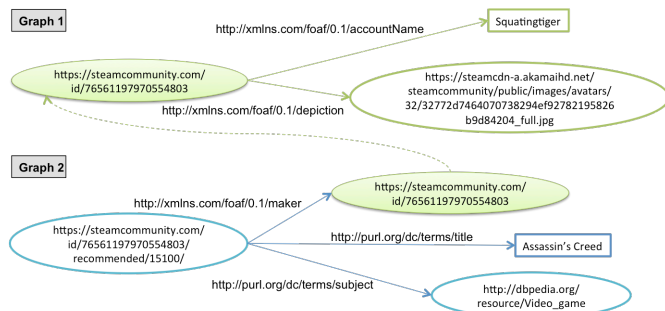[2]RDF — http://www.w3.org/TR/REC-rdf-syntax/

Fig. 1. Examples of graphs that interlink variant resources

and object are the nodes and the predicate connects the subject to the object node. The set of triples describing a particular resource form an RDF graph (Fig. 1).

RDF data can be queried by using an RDF query language called SPARQL[3]. SPARQL queries take the form of a set of triple patterns called a basic graph pattern. SPARQL triple patterns are similar to RDF triples with the difference that in a SPARQL triple, each subject, predicate and object can be bound to a variable; the variable's value to be found in the original data graph. When executing a SPARQL query, the resulting RDF data matches to the SPARQL graph pattern.

Moreover, the RDF data may require more meaning to describe its structure and therefore, an RDF vocabulary modelled using the RDF Schema (RDFS)[4] can be used to describe the RDF data's structure. Apart from vocabularies, RDF data may pertain to a specific domain which its structure needs to be explicitly defined using ontologies modelled by RDFS and/or OWL 2[5]. For example, ontologies may describe people such as the Friend of a Friend (FOAF)[6] ontology or may describe information from gaming communities to interlink different online communities such as the Semantically-Interlinked Online Communities (SIOC)[7] ontology.

C. Linked Data

As mentioned previously, when describing a particular resource within a graph, a URI is assigned to that resource which can be referred to in other graphs using that particular URI. For instance, if a particular resource represents a person within another graph that describes information about that person, the person's (resource) URI can be used for example when describing that s/he is the creator of a game review which is described in another graph; as illustrated in Fig. 1. Hence this makes it easy to link data together from different datasets and thus creating *Linked Data*[8]. Datasets which are easily accessible are linked forming the Linking Open Data (LOD) cloud[9] which forms part of the *Web of Data*. In order to publish

data in the LOD cloud, it must be structured adhering to the Linked Data principles as stated in [19] and the Data on the Web best practices as stated in [14].

The benefit of linking data is that links amongst data are explicit and try to minimise redundant data as much as possible. Therefore, similar to hyperlinks in the conventional Web that connect documents in a single global information space, Linked Data enables data to be linked from different datasources to form a single global data space [19].

IV. APPROACH

Given the core aims of this work our approach for semantic-based game generation is operationalised via the following sequence of key processes (as illustrated in Fig. 2): (1) the user inputs keywords or a narrative of the game s/he desires; (2) a list of game genres are extracted from the user's input; (3) rankings for each game are extracted from different game review and scoring sites, and the games are ranked according to the aggregate score; (4) game content, such as game plot and game walkthroughs are extracted from diverse Web sources for each of the high ranked games; (5) game information related to the different game facets (for example entities, actions etc.) is extracted from the game content and game facets RDF graphs are created; (6) the new game is generated by merging the game facets RDF graphs of the different games.

In the following subsections we detail the processes listed above and suggest ways to realise them.

A. Extracting Game Genres and Game Lists

Video game genres and lists of games for particular genres can easily be extracted from Wikidata[10] and DBpedia[11] through their SPARQL endpoints. Wikidata consists of a collaborative editing knowledge base that provides common source of data for Wikipedia[12] and it collects data in a structured form allowing data to be easily reused. DBpedia also extracts structured information from Wikipedia and publishes this structured information on the Web. Hence, both Wikidata and DBpedia are good sources of structured knowledge to extract game information already enriched in semantic meta-formats.

In our approach, a user enters some keywords that describe the game to be generated. Keyphrase extraction techniques, such as those described in [40], can be used to identify game genres from the user's input. A graph of game genres can be constructed from e.g. Wikidata, DBpedia and WordNet [15] that will be used as a reference video game genre vocabulary whilst extracting keyphrases. Once the game genre is identified, a list of games for that particular genre can be extracted from Wikidata and DBpedia. For example, the query in Fig. 3 extracts the list of *action-adventure video games* from Wikidata as illustrated in Fig. 2, where the property `P136` refers to the property `genre` and the item `Q343568` represents the `action-adventure game` genre.

[3]SPARQL — http://www.w3.org/TR/rdf-sparql-query/

[4]RDFS — http://www.w3.org/TR/rdf-schema/

[5]OWL 2 — http://www.w3.org/TR/owl2-overview/

[6]FOAF — http://www.foaf-project.org/

[7]SIOC — http://sioc-project.org/

[8]Linked Data — http://linkeddata.org/

[9]Linking Open Data (LOD) cloud — http://lod-cloud.net

[10]Wikidata — https://www.wikidata.org

[11]DBpedia — http://wiki.dbpedia.org

[12]Wikipedia — https://www.wikipedia.org

Fig. 2. An semantic approach for game generation

```
SELECT ?game ?gameLabel
WHERE
{
    ?game ?p ?o .
    ?o  ps:P136 wd:Q343568.
  SERVICE wikibase:label {
   bd:serviceParam wikibase:language "en".
   }
}
ORDER BY ?gameLabel
```

Fig. 3. A SPARQL query that extracts a list of action adventure video games from Wikidata

We envisage to extend our approach by allowing the user to input a full narrative or a detailed description of the new game. The text is then parsed to understand the narrative (relating words to their semantic meaning) by using knowledge-bases such as *GluNet* [23], that provide a semantically rich lexical commonsense vocabulary intended to assist computational storytelling in computer games. GluNet maps existing resources together from *WordNet* — a lexical database of words [15] — *VerbNet* — a lexical database of verb semantics and syntax [22] — *FrameNet* — a lexical database of frame semantics [1] — and *ConceptNet* — a commonsense knowledge-base for relating narrative concepts to games [25]. Moreover, an RDF graph can be constructed out of the semantic relationships amongst the words in the narrative, and DBpedia Spotlight [12] will be used to automatically annotate the words to their semantic representations so that they are included in the RDF graph. This RDF graph can be used when generating the new game as a reference to what the user desires.

B. Extracting Game Rankings

Game rankings can be extracted from various review sites in order to rank each game in the list extracted during the previous stage. Metacritic[13] and GameRankings[14] are two sites that provide weighted scores for games which can be used for this step of the game generation process.

Metacritic's metascores, for instance, are weighted average scores aggregated from the reviews created within the site. Metacritic game pages can easily be accessed by following the link provided in the page about the same game in Wikidata. Metacritic uses microdata to markup its pages. Microdata is a standard specification to add metadata within existing content on web pages by annotating HTML elements with machine-readable tags. The most popular vocabulary used to markup microdata content currently is Schema.org[15] that provides a collection of commonly used markup vocabularies. Therefore, the scores can easily be extracted from Metacritic by using a microdata parser to retrieve values annotated with the property `ratingValue`. Apart from scores, more semantically annotated game content can be parsed for later consumption. DBpedia also provides semantically annotated weighted scores

from various review sites such as from GameSpot[16], IGN[17], GamesRadar[18], amongst others. While the GameRankings scores are weighted average ratings from various offline and online sources currently no direct link exists to the specific GameRankings page for a particular game and a crawler would be required to find the specific page. Moreover, GameRankings does not markup its content and a scrapper would be required to extract rankings and game information.

All aforementioned scores extracted from the various sites can be defined in RDF using vocabularies such as Schema.org, and the RDF graphs would be stored in a public-accessible RDF store. These scores can, in turn, be used to rank the games in the list extracted during the previous phase.

C. Extracting Game Information

Game content, such as game plot, gameplay, game character information etc. can be found in various sources scattered around the Web, and when aggregated together, can provide in-depth details about the game mechanics. For instance Wikia[19] consists of encyclopaedias, each one specialised in a particular topic that covers much greater information and more comprehensible detail than Wikipedia. Wikia provides a detailed API that allows easy access to searching and extracting most of the content. GameFaqs[20] is another source that provides game content such as walkthroughs from which in-depth detail about level design, game design and gameplay for a particular game could be extracted. Unlike Wikia, GameFaq does not provide an API and requires Web scrapping in order to extract the content. Other sources of game information from which game mechanics could be derived from include user review sites. Most of these sources contain unstructured information and require natural language processing techniques in order to parse and process the text meaningfully.

The challenges that game information extraction brings about include: (1) how to parse and understand which text is suitable to model visual information, audio information, narrative, level design, game design or gameplay; and (2) how to semantically represent each different game facet as RDF graphs. Entity recognition techniques could be used to identify and classify entities such as persons or locations in the text. The Stanford Named Entity Recogniser (NER) [16] tool could be used to extract entities and DBpedia spotlight would be used to match these entities to DBpedia resources described in RDF. Extracting entities and relevant information about these entities could be used to describe the visual aspects of the game, game character information, or other game entities. Part-of-speech tagging could then be used to identify which words are nouns or verbs where verbs could identify what actions can be performed in a game or the rules the game would have. As explained previously, GluNet, that contains VerbNet the lexical database of verb semantics, could be used to identify

---

[13]Metacritic – http://www.metacritic.com

[14]Game Rankings — http://www.gamerankings.com/

[15]Schema.org — http://schema.org/

[16]GameSpot — http://www.gamespot.com

[17]IGN — http://www.ign.com

[18]GamesRadar — http://www.gamesradar.com/

[19]Wikia — http://www.wikia.com

[20]http://www.gamefaqs.com/

the verbs in the text. Moreover, ConceptNet could be used to identify gameplay rules; for instance in [29], the authors used both ConceptNet and WordNet to generate and define game rules. However, in [29] game rules are generated from a set of predefined verbs for a particular type of game, whereas in our work we propose that rules are generated from verbs extracted from Web content of different games (and based on common-sense reasoning using ConceptNet), creating new unconventional and unexpected game rules.

Once all game content has been parsed and classified, gaming information can be structured in RDF using common vocabularies such as SKOS[21] for describing knowledge organization systems (such as thesauri) concepts, Dublin Core[22] for describing provenance information, FOAF[23] for describing information about people, SIOC[24] for interlinking different online communities together, and Review vocabulary[25] for describing review information, amongst other vocabularies. However, new ontologies, that describe game levels or game design would be required to define new concepts which are not found in current semantic vocabularies. Game rules can be defined using the Semantic Web Rule Language (SWRL)[26] that enables Horn-like rules to be combined with an OWL knowledge-base. Ultimately, a number of RDF graphs representing various game content types can be created for each game creating an RDF store which yields a semantic knowledge-base of game information.

D. Generating Games from Semantic Information

In the final phase of the envisaged process the new game is generated by combing the several RDF graphs together, or even with other RDF graphs that are already stored in the knowledge base and that relate to the user's requirements. Ontology alignment and semantic matching techniques are used to find semantic relationships amongst ontologies [13] and to identify whether graph structures are semantically related [17] — to find the correspondences and mappings amongst the RDF graphs describing the game information. Through these correspondences, the RDF graphs are merged to form the new game graph. This new game could result to an entirely new and unconventional game genre. A semantic game engine, based on RDF and SWRL reasoners, would then be required to parse the merged RDF graphs in order to transform the semantic information into a playable game.

In its simplest instance, the ontology can directly guide the generative process — for instance an entity of type *Person* can instantiate an NPC (with details or visual appearance of the NPC provided within the ontology) — similar to the work in [2]. Moreover, the ontology can indirectly specify the parameter vector of the generator: for instance a level with the

type of "dungeon" can adjust the parameters of the generator to create levels with low linearity [34], due to the relationship between "dungeon" and "maze". A more ambitious target for semantic game generation is to use machine learning to identify patterns (e.g. visual patterns of textures, playtrace patterns in levels) of content in existing game within the ontology; the learned model of content quality and semantic information can then be used as an objective for a search-based content generator, targeting content with similar patterns as to those in existing games.

V. CONCLUSION

In this paper we presented our envisaged approach for generating games via semantic information extracted from diverse Web content. We have provided some first insights on what game content could be extracted to generate unconventional games, how to semantically enrich such content, and how games can be generated from these semantic representations. Apart from generating games, the benefit of adding semantic information to game content is many-fold: enriched game information can be published on the Web for interlinking and consumption, enriched game content can easily be reused in games without requiring any effort to modify game artefacts, and real-time interactions amongst objects in games could easily be achieved.

In contrast to current automated game generation processes such as traditional procedural content generation practices, our approach enables the use of massive amounts and dissimilar types of content from online sources. This allows content to be automatically generated whilst taking into consideration player models derived from user information stored across various online datasets [41] thereby realising a semantically-enriched version of the experience-driven PCG framework [42]. For example, Metacritic contains user reviews which can provide a quantifiable (based on the user scores) or qualitative (based on sentiment-word analysis of the textual review) model of the contributing user base. This model can be used to create game content or complete games, which are expected to appeal to the entire community or to specific parts of the community, based for instance on demographics or skill or interests collected from user's steam achievements or favoured games, respectively. On the other hand, an indirect model of player engagement with specific types of content can be gleaned from the mostly user-generated wikia pages. Pages with popular characters and locations or challenging game levels are expected to have more textual contributions (due to being updated more often by more people). This can be used to create content similar to existing game content popular in one or more wikia user communities.

With the novel approach proposed in this paper we envisage not only the generation of personalised digital games autonomously but also the creation of games that are perceived as being unconventional and unexpected, yet engaging and playable.

REFERENCES

[1] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet Project. In *17th International Conference on Computational Linguistics*, COLING'98, 1998.

[2] G. A. Barros, A. Liapis, and J. Togelius. Playing with Data: Procedural Generation of Adventures from Open Data. In *1st International Joint Conference of DiGRA and FDG*, DiGRA-FDG'16, 2016.

[3] G. A. Barros, A. Liapis, and J. Togelius. Who Killed Justin Bieber? Murder Mystery Generation from Open Data. In *Seventh International Conference on Computational Creativity*, ICCC'16, 2016.

[4] T. Berners-Lee. *Semantic Web Road Map*, September 1998.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284:34–43, 2001.

[6] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee. Linked Data on the Web (LDOW2008). In *17th International Conference on World Wide Web*, WWW '08, 2008.

[7] C. Browne and F. Maire. Evolutionary Game Design. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(1):1–16, 2010.

[8] J. T. C. Chan and W. Y. F. Yuen. Digital Game Ontology: Semantic Web Approach on Enhancing Game Studies. In *9th International Conference on Computer-Aided Industrial Design and Conceptual Design*, CAID/CD 2008, 2008.

[9] M. Cook and S. Colton. A Rogue Dream: Automatically Generating Meaningful Content For Games. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

[10] M. Cook, S. Colton, and J. Gow. Automating Game Design in Three Dimensions. *AISB Symposium on AI and Games*, 2014.

[11] M. Cook, S. Colton, and A. Pease. Aesthetic Considerations for Automated Platformer Design. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*, AIIDE'12, 2012.

[12] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes. Improving Efficiency and Accuracy in Multilingual Entity Extraction. In *9th International Conference on Semantic Systems (I-Semantics)*, 2013.

[13] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. *Ontology Matching: A Machine Learning Approach. Handbook on Ontologies*. Springer Berlin Heidelberg, 2004.

[14] B. Farias Lscio, C. Burle, and N. Calegari. W3C. Data on the Web Best Practices. 19 May 2016. W3C Working Draft. http://www.w3.org/TR/dwbp/.

[15] C. Fellbaum, editor. *WordNet: an electronic lexical database*. MIT Press, 1998.

[16] J. R. Finkel, T. Grenager, and C. Manning. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, 2005.

[17] F. Giunchiglia, P. Shvaiko, and M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. In *First European Semantic Web Symposium*, ESWS'04, 2004.

[18] K. Hartsook, A. Zook, S. Das, and M. O. Riedl. Toward supporting stories with procedurally generated game worlds. In *IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2011.

[19] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan and Claypool, 2011.

[20] A. Howlett, S. Colton, and C. Browne. Evolving pixel shaders for the prototype video game subversion. In *The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'10)*, 2010.

[21] J. Kessing, T. Tutenel, and R. Bidarra. Designing semantic game worlds. In *The Third Workshop on Procedural Content Generation in Games*, PCG'12. ACM, 2012.

[22] K. Kipper, A. Korhonen, N. Ryant, and M. Palmer. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40, 2008.

[23] B. Kybartas and R. Bidarra. A Semantic Foundation for Mixed-Initiative Computational Storytelling. In *Interactive Storytelling: 8th International Conference on Interactive Digital Storytelling*, ICIDS'15. Springer, 2015.

[24] A. Liapis, G. N. Yannakakis, and J. Togelius. Computational game creativity. In *Fifth International Conference on Computational Creativity*, ICCC'14, 2014.

[25] H. Liu and P. Singh. ConceptNet – A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*, 22(4):211–226, Oct. 2004.

[26] P. Lopes, A. Liapis, and G. N. Yannakakis. Sonancia: Sonification of procedurally generated game levels. In *1st Computational Creativity and Games Workshop*, 2015.

[27] P. Lopes, A. Liapis, and G. N. Yannakakis. Targeting Horror via Level and Soundscape Generation. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, AIIDE'15, 2015.

[28] R. Lopes and R. Bidarra. A semantic generation framework for enabling adaptive game worlds. In *8th International Conference on Advances in Computer Entertainment Technology*. ACM, 2011.

[29] M. J. Nelson and M. Mateas. Towards automated game design. In *AI* IA 2007: Artificial Intelligence and Human-Oriented Computing*, pages 626–637. Springer, 2007.

[30] K. Orland. Ars Technica: How NASA topography data brought dose of reality to SSX snowboarding courses.

[31] J. Parkkila, F. Radulovic, D. Garijo, M. Poveda-Villalón, J. Ikonen, J. Porras, and A. Gómez-Pérez. An ontology for videogame interoperability. *Multimedia Tools and Applications*, pages 1–20, 2016.

[32] J. Robertson and R. M. Young. Automated gameplay generation from declarative world representations. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, AIIDE'15, 2015.

[33] O. Sacco, M. Dabrowski, and J. G. Breslin. Linking in-game events and entities to social data on the web. In *Games Innovation Conference (IGIC), 2012 IEEE International*, pages 1–4, Sept 2012.

[34] G. Smith and J. Whitehead. Analyzing the Expressive Range of a Level Generator. In *Workshop on Procedural Content Generation in Games*, PCGames'10. ACM, 2010.

[35] J. Togelius, G. N. Yannakakis, K. O. Stanley, and C. Browne. Search-Based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):172–186, Sept 2011.

[36] M. Treanor, B. Blackford, M. Mateas, and I. Bogost. Game-O-Matic: Generating Videogames That Represent Ideas. In *The Third Workshop on Procedural Content Generation in Games*, PCG'12. ACM, 2012.

[37] T. Tutenel, R. Bidarra, R. M. Smelik, and K. J. D. Kraker. The Role of Semantics in Games and Simulations. *Computers in Entertainment*, 6(4):57:1–57:35, Dec. 2008.

[38] T. Tutenel, R. M. Smelik, R. Bidarra, and K. J. de Kraker. Using Semantics to Improve the Design of Game Worlds. In *5th Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE'09, 2009.

[39] I. D. Visualization. Speedtree., 2010.

[40] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning. Kea: Practical automatic keyphrase extraction. In *The fourth ACM conference on Digital libraries*. ACM, 1999.

[41] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. Andre. Player Modeling. In *S. M. Lucas, et al. eds. Artificial and Computational Intelligence in Games*. s.l.: Dagstuhl Seminar, 2013.

[42] G. N. Yannakakis and J. Togelius. Experience-driven procedural content generation (extended abstract). In *International Conference on Affective Computing and Intelligent Interaction*, ACII'15, 2015.

[43] J. P. Zagal and A. Bruckman. The Game Ontology Project: Supporting Learning While Contributing Authentically to Game Studies. In *8th International Conference on International Conference for the Learning Sciences*, ICLS'08, 2008.