Extracting Semantic-Based Video Game Characters Information from Social Media Platforms

Owen Sacco*, Antonios Liapis, Georgios N. Yannakakis

Institute of Digital Games, University of Malta, Msida, Malta

Email address: owensacco@gmail.com (O. Sacco), antonios.liapis@um.edu.mt (A. Liapis), georgios.yannakakis@um.edu.mt (G. N. Yannakakis)

*Corresponding author

To cite this article:
Owen Sacco, Antonios Liapis, Georgios N. Yannakakis. Extracting Semantic-Based Video Game Characters Information from Social Media Platforms. *Mathematics and Computer Science*. Vol. 4, No. 1, 2019, pp. 24-39. doi: 10.11648/j.mcs.20190401.13

Received: March 17, 2019; Accepted: April 30, 2019; Published: May 23, 2019

Abstract: Character generation in video games currently relies on game developers manually creating game characters which costs in time, effort and resources. Social media, in the form of blogs, microblogs, forums, wikis, social networks and review sites contain rich information about characters in video games that are not exploited for character generation. However, such information contained in various social media applications are disconnected from one another and are not structured or enriched that can be utilised for character generation. Semantic Web techniques provide ways of linking and enriching information contained in disconnected datasets. This enriched information can be used to build complete character models for generating new characters in video games. Moreover, a video game character knowledge graph can be constructed out of the semantically-enriched information that can be used not only for character generation in video games, but also in any application that requires information about video game characters. In this paper, we present our approach for exploiting social media platforms to create semantically-enriched character models. In particular, we present our Game Character Ontology (GCO) – a light-weight vocabulary for describing character information in video games – and our methodology for extracting and describing (using our ontology) game character information from social media platforms.

Keywords: Vocabularies, Ontologies, Semantic Web, Computer Games Technology, Procedural Content Generation

1. Introduction

Social media platforms, consisting of wiki-based systems, social networks, review sites, blog sites, and microblog sites (amongst others), provide users with systems to create their own content and this resulted in the large amount of content currently available on the Web. Several of these platforms are specific to the creation of content related to information about video games, in particular, these social media platforms contain rich information about video game characters. The information in these datasets includes personal attributes such as name or age, relationship information with other characters, personality and characteristics information, biographical information, equipment and skills information, locations which the character has visited or lived in, reviews and rankings about a character from users who have played with the character, etc. However, the content in these diverse platforms, although they relate to the same game, in our case the same game character, the information is currently disconnected from one another. These vast sources of game content could be reused to reduce the development time and effort to create games. Our vision is to generate novel and semantically-enriched content for games from diverse Web sources [38]. In particular, we envision a game character generator that extracts character information directly from the Web such as from wiki articles or images that are freely available, and generate new characters from already existing ones. 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.

Games are composed of different domains (or facets) that contribute to the game's look, feel and experience [28]. These facets include visuals, audio, narrative, gameplay, game rules and game levels. 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 [41] that generates trees and

vegetation as part of the visuals facet, or the Ludi system [19]

which generates game rules for two-player board games as

part of the game rules facet. Although such algorithms are

beneficial for automatic generation of content, it is still rare

that game characters are considered.

The datasets created by social media communities contain

information which can be used to generate or reuse game

character content in games, but are not easily discoverable.

The emerging Web of Data trend [15], where datasets are

published in a standard form for easy interlinking, enables to

especially view the whole Web as one massive integrated

database. Nevertheless, game character 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 game character content, this would

enable Web content to be reused in games. Moreover, the

representation of semantically-enriched and semantically-

interlinked game character content would enable game

character generators or game character asset managers to infer

how characters 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 present our framework to extract video

game character content from social media platforms, such as

from wiki-based systems -- e.g. Fandom [4] and Giant Bomb

[5]. The extracted character content is semantically

annotated using our Game Character Ontology (GCO) [36] --

a light-weight vocabulary for describing character

information in video games. The remainder of this paper is as

follows: Section 2 provides use case scenarios for using

extracted character information. Section 3 reviews current

work on semantics in games, existing game ontologies for

describing character information, and methodologies for

extracting and semantic-annotating text. Section 4 offers

core background information about the Web of data. In

Section 5 we discuss the type of game content available in

social media platforms from which character information can

be extracted. This section also provides an overview of our

Game Character Ontology (GCO) and provides some

examples for describing game character content using GCO.

Section 6 provides our methodology for extracting and

semantically annotating content from social media platforms

using GCO, and section 7 provides preliminary results of the

extraction methodology. Section 8 concludes the paper by

providing an overall discussion about the future steps of our

work.

2. Motivations

As mentioned in section 1, current social media platforms

provide rich video game content which are not exploited for

game content generation. Since games are complex and

contain many types of content, in this paper we focus

specifically on content describing game characters. Suppose a

framework that can automatically extract game character

information from various social media platforms, and this

content is inter-linked and semantically-annotated to provide

comprehensive character models. These models would be

described in RDF (see Section 4 for a detailed description of

RDF) using the Game Character ontology (GCO) and would

collectively create a game character dataset. This dataset

would contain rich information about characters that can be

reused in games and/or for generating new unconventional

game characters from existing ones. This game character

dataset would also provide character assets for generating

characters using game designer applications. For example, if a

designer requires a Mage in a particular game which s/he is

designing, the designer could reuse the details of mage

characters already existing in a game character dataset without

having to re-create a mage from scratch. This would reduce

the time and cost for generating video game characters in

video games. Moreover, games could be designed to

automatically generate (playable or non-playable) characters

during gameplay from such game character dataset without

the need of pre-creating characters in games. Furthermore,

several games, such as the Baldur's Gate series [2],

Neverwinter Nights series [6] and The Elders Scrolls series [8],

provide users an in-game character generator to create their

character (which they will play with) from a pre-determined

set of character information manually created by developers.

We envisage that in-game character generators will exploit a

semantic-based game character dataset which would provide

more character details from which users can choose from

without requiring developers to manually create all this

content. This will provide users to create unconventional

characters during gameplay.

3. Related Work

Semantics in games is still in its infancy and perhaps the

closest attempt at using structured real-world data in games is

the Data Adventures project [11, 12] which uses SPARQL

queries (see Section 4 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 advantages of using rich semantic information to

automatically generate games are numerous [39] 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 [27,

29, 40]. Although these provide useful insights in generic

semantic models that describe interactions with game worlds,

they do not offer vocabularies for describing game content
such as game characters 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 some key game-based ontologies currently existent. The Game Ontology Project [45] is a wiki-based knowledge-base that aims to provide elements of gameplay. However, this project does not take into consideration character games. Moreover, it does not provide a vocabulary to be consumed by data described in RDF which could make it potentially useful for game character generation. The Digital Game Ontology [21] provides a general game 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. The Ludo ontology [35] provides concepts that describe different aspects of serious games, however, it does not provide detailed concepts for describing game characters. The SALERO virtual character ontology [17] provides a generic ontology for describing characters in media production. Although it provides a generic model, it does not provide detailed concepts for defining game character information such as the species, race, character class, skills, weapons etc. The authors in [23] provide a generic ontology for defining RPG games and do not provide detailed concepts to describe game characters for other genre types. Finally, the Video Game Ontology (VGO) [33] provides concepts for defining interoperability amongst video games and the Game2Web ontology [37] 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 and hence do not provide features for describing detailed game character information.

Our ontology was created since although similar classes and properties could be found in other ontologies, these do not imply that they are properties of fictional characters. For example, most properties found in the FOAF vocabulary [18] such as foaf:name are normally used for describing real persons who exist or have existed, and therefore, this vocabulary cannot be used to describe fictional game character information. Another example would be the classes and properties defined in DBpedia's ontology [3] such as dbo:skinColor, dbo:hairColor or dbo:eyeColor that describe physical features of real persons specified by the domain dbo:Person implying that the subject is a real person. Moreover, the Appearance Ontology [42] provides classes and properties for defining appearance and gender features of real persons specified by the domain appearances:Person implying that the subject is a real person. Therefore, these classes and properties cannot be used to describe properties of fictional characters, and new classes and properties are required.

Content extraction methods are also relevant to our approach; hence we outline some existing work about information extraction. GATE, an open-source platform, provides several tools for natural language processing [16] and it has been used in a significant amount of projects, for example in [32] and [22]. Due to its popularity, our methodology for information extraction will be using the GATE platform. Several work exist on semantic annotation of documents, such as [20], [26] and [34]. Our methodology is similar to the work in [34] and we demonstrate in this paper how this methodology can be used for semantically extracting game character information.

4. 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.

4.1. The Semantic Web

The Semantic Web [14] 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 [13] is the Resource Description Framework (RDF)\(^1\).

4.2. 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

---

1 RDF -- http://www.w3.org/TR/REC-rdf-syntax/
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 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 (Figure 1).

**Figure 1.** Examples of graphs that interlink variant resources.

**Figure 2.** Article about Altair Ibn-La’Ahad on Giant Bomb.
RDF data can be queried by using an RDF query language called SPARQL\(^2\). 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)\(^3\) 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\(^4\). For example, ontologies may describe people such as the Friend of a Friend (FOAF)\(^5\) ontology or may describe information from gaming communities to interlink different online communities such as the Semantically-Interlinked Online Communities (SIOC)\(^6\).

4.3. 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

---

\(^2\) SPARQL — http://www.w3.org/TR/rdf-sparql-query
\(^3\) RDFS — http://www.w3.org/TR/rdf-schema
\(^4\) OWL 2 — http://www.w3.org/TR/owl2-overview
\(^5\) FOAF — http://www.foaf-project.org
\(^6\) SIOC — http://sioc-project.org/ontology
when describing that s/he is the creator of a game review which is described in another graph; as illustrated in Figure 1. Hence this makes it easy to link data together from different datasets and thus creating Linked Data. Datasets which are easily accessible are linked forming the Linking Open Data (LOD) cloud 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 [25] and the Data on the Web best practices as stated in [24].

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 [25].

5. Game Character Models

Game character content is available in wiki-based sites such as Fandom [4] and Giant Bomb [5]. For example, Figure 2 and Figure 3 illustrate game character articles from both platforms respectively about the same character found in the Assassin's Creed series [1] named Altair Ibn-La'Ahad. As can be noted, such articles provide comprehensive detail about characters from which character models can be created. Other content can be found in user-based review sites such as reviews in Steam Discussions platform [7]. Most of the content found in social media systems is unstructured and thus makes it hard to be processed and reused in games. Therefore, this content must be transformed into structured content and enriched with semantic-annotations in order to add more meaning to the character's attributes. In this section, we provide an overview of the Game Character Ontology (GCO) [36] -- an ontology for describing game character models from content extracted from social media platforms. This section also provides an example of the information about Altair Ibn-La’Ahad defined using GCO.

5.1. Game Character Ontology (GCO)

The Game Character Ontology (GCO) [36] (illustrated in Figure 4) provides a light-weight vocabulary for describing characters in video games. Information about video game characters can be extracted from various social media platforms and can be described using this ontology to create semantic video game character models in the form of graphs. These models can be reused in games or can be used to generate new game characters as a result of merging and combining various game character graphs.

This ontology provides the following classes, each of which provide several properties that describe various detail information about characters:

1. Character: this is the main character class that defines the entity being a game character. This class provides several properties that describe various generic personal details of a character, for example the character's name, age, date of birth, birth place, etc. Game characters can be defined as either playable or non-playable, depending on whether the character can be controlled or not by a player.

2. Appearance: this class provides various properties for defining the physical appearance of the character, for example the character's facial hair colour and/or skin colour. It also describes the various parts of the character's physical body.

3. Personality: describes various aspects of the character's personality for example whether the character is an angry (hot-tempered) and/or arrogant character.

4. Gender: defines whether the character is male or female.

5. Species: represents a generic classification of a character for example Human or Elf species.

6. Race: represents a more specific classification than the species classification for example Man or Woman or Moon Elf race.

7. Character's Class: represents a more specific classification based on archetypes and careers for example the character class Assassin. A character could be defined with more than one class, for example a character can be assigned to a Mage class and Fighter class.

8. Role: defines specific representations of roles that characters can change into without changing the character class or occupation. For example, a character could change into a Thief role, having different skills, abilities, outfit, weapon, armour etc., without changing the character's classes of a Mage and Fighter class.

9. Skill defines active actions (activated by a player) which characters can perform to complete a specific task in a game, for example running, climbing and grabbing.

10. Ability defines passive actions that are applied automatically and are not activated by a player, for example “increase the character's armour by 20%” or “enhance sniper rifle hit points by 10 once a specific ability is obtained”.

11. Power defines unique actions that are activated by a player which normally consist of super-natural actions in nature for example magic spells or sources of energy that power up armour.

12. Relationship: defines different relationship types which a character has with other characters for example whether a character is an enemy of another character.

13. Outfit: defines various garments that characters could wear on different parts of their body.

14. Weapon: defines various attributes of weapons a character could use to cause damage to opponents (or other characters). Weapons are also bound to actions described by the Action class.

7 Linked Data -- http://linkeddata.org
8 Linking Open Data (LOD) cloud -- http://lod-cloud.net
15. Armour: defines various pieces of armour a character could use or wear to deflect any damage caused by opponents.
16. Inventory: defines a set of items which characters can use or carry, for example health potions.
17. Group: defines various types of groups in which characters could be a part of -- from formal groups such as the Brotherhood of Assassins or the Templar Order in the Assassin's Creed series [1] to informal groups where different characters team up to solve a common goal such as in the Uncharted series [10] or the Tomb Raider series [9].

5.2. Defining Game Character Information with (GCO)

The following figures illustrate an example of a game character's information described using GCO. The information was extracted from the articles illustrated in Figure 2 and Figure 3 which contain information about the character named Altair Ibn-La’Ahad. This example contains:
1. Figure 5 illustrates the general details of the character such as name and date of birth. The age of the character can be calculated during the game by finding the difference between the game's current date and the character's birth date.
2. Figure 6 illustrates the character's personality.
3. Figure 7 illustrates the character's species, race and character class.
4. Figure 8 illustrates the character's skills.
5. Figure 9 illustrates the character's preferred garments.
6. Figure 10 illustrates the character's preferred armour and weapons.
7. Figure 11 illustrates the character's affiliations with groups.

![Figure 4. Overview of the Game Character Ontology (GCO).](image)
character:Altair_ibn_La-Ahad
  a gco:PlayableCharacter;
  gco:name "Alta\"i\r Ibn-La\’Ahad";
  gco:birthDate "1165-01-11"^^xsd:dateTime;
  gco:birthYear "1165"^^xsd:year;
  gco:birthPlace location:Masyaf;
  gco:deathDate "1257-08-12"^^xsd:dateTime;
  gco:deathYear "1257"^^xsd:year;
  gco:deathPlace location:Masyaf;
  gco:hasGender gco:Male;
  gco:hasImage character:Altair_ibn_La-Ahad.png;
  gco:currentOccupation occupation:Mentor;
  gco:currentOccupation occupation:Master_Assassin;
  gco:description "Altair Ibn-La’Ahad (1165 – 1257) was a Syrian-born member of the Levantine Brotherhood of Assassins and served as their Mentor from 1191 until his death in 1257. During his tenure as Mentor, Altair made several discoveries ...";
  gco:appearsIn game:Assassin%27s_Creed:_Altair%27s_Chronicles;
  gco:appearsIn game:Assassin%27s_Creed;
  gco:appearsIn game:Assassin%27s_Creed:_Bloodlines;
  gco:appearsIn game:Assassin%27s_Creed:_IX;
  gco:appearsIn game:Assassin%27s_Creed:_Revelations;
  gco:appearsIn game:Assassin%27s_Creed:_Memories;
  gco:voicedBy actor:Philip_Shahbaz;
  gco:voicedBy actor:Cas_Anvar;
  gco:onlineArticle <http://assassinscreed.wikia.com/wiki/Alta%0C3%A5Fr_Ibn-La\'Ahad>;
  gco:onlineArticle <http://www.giantbomb.com/altair-ibn-laahad/3005-48/>;

[...]

Figure 5. Altair Ibn-La’Ahad's General Details.
Figure 6. Altair Ibn-La'Ahad's Personality.

Figure 7. Altair Ibn-La'Ahad's Species, Race and Character Class.

Figure 8. Altair Ibn-La'Ahad's Skill.

Figure 9. Altair Ibn-La'Ahad's Outfit.
Figure 10. Altair Ibn-La'Ahad's Armour and Weapons.

Figure 11. Altair Ibn-La'Ahad's Group Affiliations.

Figure 12. Altair Ibn-La'Ahad's Relationship Information.
Figure 12 illustrates relationship information which the character has with other characters. All the relationship information contained within a game character dataset create a network of relationships amongst characters, similar to relationships amongst friends found in a social network. Hence, this creates a game character network.

Figure 13. A SPARQL query that extracts a list of action adventure video game titles from Wikidata.

Figure 14. A SPARQL query that extracts a list of video game characters in the Assassin's Creed game from Wikidata.

6. Methodology: Semantic-Based Extraction of Game Character Content

Our approach for semantic-based game character extraction contains the following sequence of processes:

1. A list of game titles is extracted for a particular game genre -- in this paper we focus on action-adventure games;
2. For each game title, a list of characters is extracted; and
3. For each character, information is extracted from the online content and are described as RDF graphs using GCO.

In the following subsections we detail the processes listed above.

6.1. Extracting Game Titles and Characters List

Video game titles for particular genres can easily be extracted from Wikidata and DBpedia through their SPARQL endpoints. Wikidata consists of a collaborative editing knowledge base that provides common source of data for Wikipedia 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. For example, the query in Figure 13 extracts the list of action-adventure video games from Wikidata, where the property P136 refers to the property genre and the item Q343568 represents the action-adventure game genre. Table 1 presents a snippet of game titles extracted from Wikidata using this query.

Table 1. Action-Adventure Game Titles.

| Game URI   | Game Title               |
|------------|--------------------------|
| wd:Q420292 | Assassin’s Creed         |
| wd:Q211735 | Assassin’s Creed II      |
| wd:Q40166  | Assassin’s Creed III     |
| wd:Q76868  | Assassin’s Creed III: Liberation |
| wd:Q6052688| Assassin’s Creed IV: Black Flag |
| wd:Q54617566| Assassin’s Creed Odyssey |
| wd:Q30138024| Assassin’s Creed Origins |
| wd:Q18602166| Assassin’s Creed Syndicate |
| wd:Q677351 | Assassin’s Creed: Brotherhood |
| wd:Q739654 | Assassin’s Creed: Revelations |
| wd:Q317620 | Tomb Raider             |
| wd:Q816451 | Tomb Raider II          |
| wd:Q1123794| Tomb Raider III         |
| wd:Q580667 | Tomb Raider: Anniversary |
| wd:Q665785 | Tomb Raider: Legend     |
| wd:Q621616 | Tomb Raider: The Angel of Darkness |
| wd:Q686568 | Tomb Raider: The Last Revelation |
| wd:Q915860 | Tomb Raider: Underworld |
| wd:Q17150  | Uncharted: Drake’s Fortune |
| wd:Q17146  | Uncharted: Among Thieves |
| wd:Q17138  | Uncharted: 3: Drake’s Deception |
| wd:Q16681843| Uncharted: 4: A Thief’s End |
| wd:Q28088198| Uncharted: The Lost Legacy |

Once the game titles are extracted, for each game title a list of characters is extracted from Wikidata. For example, the query in Figure 14 extracts a list of characters that appear in the Assassin's Creed game, where the property P1441 refers to the property “present in work” and the item Q420292 represents the Assassin’s Creed game. Table 2 presents a snippet of characters that appear in the Assassin's Creed game extracted from Wikidata using this query.

Table 2. Assassin’s Creed Character List.

| Character URI | Character Name         |
|---------------|------------------------|
| wd:Q4063579   | Altair Ibn-La’Ahad     |
| wd:Q18711529  | Arno Dorian            |
| wd:Q18416899  | Aveline de Grandpre    |
| wd:Q6870802   | Connor Kenway          |
| wd:Q2035249   | Desmond Miles          |
| wd:Q6426771   | Edward Kenway          |
| wd:Q994344    | Ezio Auditore di Firenze |
| wd:Q25407606  | Haytham Kenway         |
| wd:Q26466448  | Lucy Stillman          |
| wd:Q3847575   | Maria Thorpe           |

6.2. Extracting Game Character Information

Game content such as game character information, as mentioned above, can be found in various sources scattered around the Web, and when aggregated together, can provide in-depth details about the game character mechanics. For instance, the wiki-based systems Fandom and GiantBomb...
illustrated in section 5 consist of encyclopaedias specialised in game topics that covers much greater information and more comprehensible detail than Wikipedia. Both Fandom and GiantBomb provide an API that allows easy access to searching and extracting most of the content. However, the APIs do not provide all the content contained in the articles and therefore Web scrambling is required in order to extract the content. The information extracted from these sources is unstructured 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 character information; and
2. How to semantically represent game characters as RDF graphs.

Game characters can be semantically represented using the Game Character Ontology (GCO) described in subsection 5.1. With regards to parsing and semantically-annotating the text, this involves information extraction techniques, in particular semantic annotation techniques since our aim is to annotate the text by mapping the entities to their semantic-representations described in GCO. Our semantic annotation methodology is similar to the methodology described in the work [34] where they annotate news articles based on a pre-defined general news items ontology.

Semantic annotation involves named-entity recognition (NER) techniques to identify and annotate entities such as characters, locations or dates in the text to their semantic-representations in GCO. Moreover, the relations amongst the entities form part of the attributes of an entity. In classical named-entity recognition techniques, general entity types are used (modelled based on real-world entities) to annotate domain-independent entities. However, by using general NER for game character information is not sufficient since real-world entities do not take into consideration fictional game entities. Moreover, several fictional entities may refer to real-world entities (for example fictional characters modelled based on real-world persons), however, these fictional entities are not the same as their real-world entity representations since the fictional entities could have an alternate reality containing different information from their real-world entities. Therefore, a pre-populated knowledge base of generic game character entities is used for the semantic annotation process. This knowledge base is iteratively enriched with game character entities and relations as a result of the information extraction process.

The semantic annotations are kept separately from the original wiki-content and these are stored in a semantic repository. Ontotext GraphDB database [12] is used as a semantic repository since it is based on OWL, it is industry-tested (i.e. several industrial semantic applications use this database) and it also supports RDF4J [13] - a Java framework for processing and handling RDF data.

The information extraction process in our methodology uses the General Architecture for Text Engineering (GATE) platform [14] [16]. This platform is an open source software for text processing and incorporates several plugins such as GATE’s built-in plugin called ANNIE (A Nearly-New Information Extraction System) which provides a ready-made information extraction pipeline consisting of tokenization, sentence splitting and part-of-speech (POS) tagging. The tokenizer splits the text into simple units, known as tokens, such as words, numbers, punctuation and space tokens. GATE’s tokenizer relies on a set of regular expression rules which are compiled into a finite-state machine. This is followed by a sentence splitter that splits the text into sentences by determining whether punctuation such as full stops denote the end of a sentence. Part-of-speech tagging involves assigning grammatical categories to each token based on the definition and context of the words -- i.e. identifying whether the tokens are nouns, verbs, adjectives, adverbs, etc. ANNIE also uses a gazetteer for identifying entities in text which consist of a set of lists containing names of entities. These lists are used to find occurrences of entities in text as part of the named-entity recognition process, and the gazetteer lists can be modified according to the domain of the text. ANNIE also provides a semantic tagger which contain rules to produce annotated entities. Apart from ANNIE, GATE incorporates other third-party NLP pipeline frameworks for information extraction such as Stanford CoreNLP [31] and Apache OpenNLP [16]. GATE also provides an Ontology plugin for loading ontologies in GATE and annotating text with classes and properties within the loaded ontology.

In our approach, ANNIE was used as the information extraction pipeline for tokenization, sentence splitting and POS tagging. A modified gazetteer was used to incorporate fictional entities. This was used to look up entities in the text, as part of the pattern-matching grammar NER process built in ANNIE. The Ontology plugin was used to load the GCO ontology and for annotating the entities with their semantic descriptions using GCO as described and illustrated in subsection 5.2. Therefore, given the GCO ontology, the entities extracted can be linked to their semantic representations. The semantic descriptions are stored in Ontotext’s Graph DB by using RDF4J to write and store the semantic descriptions in RDF. This dataset acts as our knowledge graph of semantic-based game character information.

7. Evaluation

GATE provides a variety of tools for evaluation, one of which is the Corpus Benchmark tool plugin that compares annotation sets over an entire document or corpus against a gold standard. Our gold standard consists of 100 human semantically annotated character information in highly-ranked action-adventure video games (namely the Assassin’s Creed Series [1], the Uncharted Series [10] and the Tomb Raider Series [9]) described in Fandom [4] and Giant Bomb [5]. Table 3 presents a snippet of the information extracted and

---

12 GraphDB -- http://ontotext.com/products/graphdb
13 RDF4J -- http://rdf4j.org
14 GATE -- https://gate.ac.uk
15 CoreNLP -- http://stanfordnlp.github.io/CoreNLP
16 OpenNLP -- https://opennlp.apache.org
Table 3. Annotations

| Value                        | Annotation Type                                      |
|------------------------------|-----------------------------------------------------|
| Altair Ibn-La`Ahad           | geo:Character / geo:PlayableCharacter / geo:name     |
| Ezio Auditore da Firenze     | geo:Character / geo:PlayableCharacter / geo:name     |
| Ratonhnhake-ton / Connor     | geo:Character / geo:PlayableCharacter / geo:name     |
| Aveline de Grandpré          | geo:Character / geo:PlayableCharacter / geo:name     |
| Edward Kenway                | geo:Character / geo:PlayableCharacter / geo:name     |
| Shay Patrick Cormac          | geo:Character / geo:PlayableCharacter / geo:name     |
| Arno Victor Dorian           | geo:Character / geo:PlayableCharacter / geo:name     |
| Jacob Frye                   | geo:Character / geo:PlayableCharacter / geo:name     |
| Evie Frye                    | geo:Character / geo:PlayableCharacter / geo:name     |
| Bayek                        | geo:Character / geo:PlayableCharacter / geo:name     |
| Lara Croft                   | geo:Character / geo:PlayableCharacter / geo:name     |
| Jacqueline Natla             | geo:Character / geo:NonPlayableCharacter / geo:name  |
| Winston Smith                | geo:Character / geo:NonPlayableCharacter / geo:name  |
| Nathan Drake                 | geo:Character / geo:PlayableCharacter / geo:name     |
| Elena Fisher                 | geo:Character / geo:NonPlayableCharacter / geo:name  |
| Victor Sullivan              | geo:Character / geo:NonPlayableCharacter / geo:name  |
| Chloe Frazer                 | geo:Character / geo:PlayableCharacter / geo:name     |
| Male                         | geo:Gender                                           |
| Female                       | geo:Gender                                           |
| Cold                         | geo:Personality                                      |
| Objective                    | geo:Personality                                      |
| Rebellious                   | geo:Personality                                      |
| Impetuous                    | geo:Personality                                      |
| Demanding                    | geo:Personality                                      |
| Arrogant                     | geo:Personality                                      |
| Impatient                    | geo:Personality                                      |
| Calm                         | geo:Personality                                      |
| Wise                         | geo:Personality                                      |
| Mentor                       | geo:Occupation                                       |
| Master Assassin              | geo:Occupation / geo:CharacterClass                  |
| Assassin                     | geo:Occupation / geo:CharacterClass / geo:Role       |
| Slave                        | geo:Occupation / geo:CharacterClass / geo:Role       |
| Medjay                       | geo:Occupation / geo:Group                           |
| Pirate                       | geo:Occupation / geo:CharacterClass                  |
| Archaeologist                | geo:Occupation / geo:CharacterClass                  |
| Treasure Hunter              | geo:Occupation / geo:CharacterClass                  |
| Masyaf                       | geo:Place                                            |
| Florence                     | geo:Place                                            |
| Kanatashseton                | geo:Place                                            |
| New Orleans                  | geo:Place                                            |
| West Indies                  | geo:Place                                            |
| New York                     | geo:Place                                            |
| Versailles                   | geo:Place                                            |
| London                       | geo:Place                                            |
| Siwa                         | geo:Place                                            |
| Lost Island                  | geo:Place                                            |
| Atlantis                     | geo:Place                                            |
| Rub` al Khali Desert         | geo:Place                                            |
| Colombia                     | geo:Place                                            |
| King’s Bay                   | geo:Place                                            |
| Madagascar                   | geo:Place                                            |
| Human                        | geo:Species                                          |
| Man                          | geo:Race                                             |
| Woman                        | geo:Race                                             |
| Lady                         | geo:Role                                             |
| Walking                      | geo:Skill                                            |
| Running                      | geo:Skill                                            |
| Climbing                     | geo:Skill                                            |
| Jumping                      | geo:Skill                                            |
| Eavesdropping                | geo:Skill                                            |
| Interrogation                | geo:Skill                                            |
| Pickpocketing                | geo:Skill                                            |
| Eagle Vision                 | geo:Skill                                            |
| Weapon Combat                | geo:Skill                                            |
| Value                                         | Annotation Type |
|----------------------------------------------|-----------------|
| Hand-to-hand Combat                         | gco:Skill       |
| Blade Throwing                              | gco:Skill       |
| Stealth                                      | gco:Skill       |
| Health Boost I – Increase your health by 1 segment | gco:Ability     |
| Health Boost II – Increase your health by 2 segments | gco:Ability     |
| Health Boost III – Increase your health by 3 segments | gco:Ability     |
| Unstoppable I – Increase the time before you lose your combo streak due to inaction | gco:Ability     |
| Unstoppable II – The first hit you take during a combo doesn’t end your combo | gco:Ability     |
| Robe                                         | gco:Outfit      |
| Sash                                         | gco:Outfit      |
| Hood                                         | gco:Outfit      |
| Collar                                       | gco:Outfit      |
| Belt                                         | gco:Outfit      |
| Boots                                        | gco:Outfit      |
| Cape                                         | gco:Outfit      |
| Trousers                                     | gco:Outfit      |
| Shirt                                        | gco:Outfit      |
| Dress                                        | gco:Outfit      |
| Shorts                                       | gco:Outfit      |
| Tank top                                     | gco:Outfit      |
| Health potion                                | gco:Inventory   |
| Trinket                                      | gco:Inventory   |
| Goblet                                       | gco:Inventory   |
| Key                                          | gco:Inventory   |
| Animal goods                                 | gco:Inventory   |
| Papyri                                       | gco:Inventory   |
| Documents                                    | gco:Inventory   |
| Sword                                        | gco:Weapon      |
| Spear                                        | gco:Weapon      |
| Mace                                         | gco:Weapon      |
| Dagger                                       | gco:Weapon      |
| Bow                                          | gco:Weapon      |
| Arrow                                        | gco:Weapon      |
| Makarov PM                                   | gco:Weapon      |
| Beretta 92FS Brigadier Inox                 | gco:Weapon      |
| Colt Anaconda                                | gco:Weapon      |
| MP 40                                        | gco:Weapon      |
| AK-47                                        | gco:Weapon      |
| Sniper Rifle                                 | gco:Weapon      |
| M79 Grenade Launcher                        | gco:Weapon      |
| Hand Grenade                                 | gco:Weapon      |
| Leather Armour                               | gco:Armour      |
| Venom Gloves                                 | gco:Armour      |
| Viper’s Hood                                 | gco:Armour      |
| Slithering Belt                              | gco:Armour      |
| Noxious Boots                                | gco:Armour      |
| Agamenmon’s Waistband                        | gco:Armour      |
| Body Armour                                  | gco:Armour      |
| Gauntlets                                    | gco:Armour      |
| Hero Belt                                    | gco:Armour      |
| Helmet                                       | gco:Armour      |
| Bulletproof Vest                             | gco:Armour      |
| Shield                                       | gco:Armour      |
| Child                                        | gco:Relationship|
| Son                                          | gco:Relationship|
| Daughter                                     | gco:Relationship|
| Brother                                      | gco:Relationship|
| Father                                       | gco:Relationship|
| Sister                                       | gco:Relationship|
| Spouse                                       | gco:Relationship|
| Husband                                      | gco:Relationship|
| Parent                                       | gco:Relationship|
| Enemy                                        | gco:Relationship|
| Colleague                                    | gco:Relationship|
| Close Friend                                 | gco:Relationship|
| Assassin Brotherhood                         | gco:Group       |
GATE provides different metrics for evaluation, and we use the precision, recall and F-measure ($F_1$ score) metric to measure the system-annotated characters against the human-annotated characters. This measure is used for our approach since this metric is widely used for evaluating information extraction methodologies [30]. Precision measures the number of correctly identified entities as a percentage of the number of entities identified, recall measures the number of correctly identified items as a percentage of the total number of correct entities, and the F-measure score is a weighted average of both precision and recall when $\beta = 1$ (i.e. both precision and recall are weighted equally). In table 4 we present the results of this measure that illustrates the precision, recall and F-measure score for the common annotations found in GCO. These results represent the correctness of the system semantically annotating game character information using classes and properties described in GCO.

| Annotation Type       | Value                                      | Annotation Type       |
|-----------------------|--------------------------------------------|-----------------------|
| Character             | Templars                                   | gco:Group             |
|                       | Edelweiss Pirates                          | gco:Group             |
|                       | Followers of Romulus                       | gco:Group             |
|                       | Freemasons                                 | gco:Group             |
|                       | Green Gang                                 | gco:Group             |
|                       | Founding Fathers                           | gco:Group             |
|                       | Ship Crew                                  | gco:Group             |
|                       | Order of the Ancients                      | gco:Group             |
|                       | Deathless Ones                             | gco:Group             |
|                       | The Immortals                              | gco:Group             |

### Table 4. Average precision, recall and $F_1$ score when $\beta = 1$.

| Annotation Type       | Precision | Recall | $F_1$ Score |
|-----------------------|-----------|--------|-------------|
| Character             | 78%       | 70%    | 74%         |
| Gender                | 71%       | 82%    | 76%         |
| Personality           | 81%       | 74%    | 77%         |
| Occupation            | 75%       | 71%    | 73%         |
| Place                 | 71%       | 75%    | 73%         |
| Species               | 65%       | 61%    | 63%         |
| Race                  | 62%       | 67%    | 64%         |
| Character Class       | 68%       | 60%    | 64%         |
| Role                  | 61%       | 60%    | 60%         |
| Skill                 | 70%       | 65%    | 67%         |
| Ability               | 60%       | 62%    | 61%         |
| Power                 | 62%       | 59%    | 60%         |
| Outfit                | 72%       | 67%    | 69%         |
| Inventory             | 66%       | 60%    | 63%         |
| Weapon                | 74%       | 78%    | 76%         |
| Armour                | 79%       | 74%    | 76%         |
| Relationship          | 62%       | 64%    | 63%         |
| Group                 | 65%       | 62%    | 63%         |

8. Conclusion

In this paper we presented our game character ontology and extraction methodology for semantically annotating game character content from diverse Web content. This semantic information can be used for automatic game character generation and also to build game character knowledge graphs. In contrast to current automated game generation processes such as traditional procedural content generation (PCG) 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 [43] thereby realising a semantically-enriched version of the experience-driven PCG framework [44]. The ontology and methodology presented in this paper can be used to generate game characters for creating video 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 fandom 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 characters similar to existing game content popular in one or more fandom user communities. Moreover, new characters can be created from already existing character models. For example, if a game designer requires to create a new character of type *assassin*, all the skills, abilities and other attributes that are already semantically-described for existing characters of the same type could be inherited by the new character without the designer having to recreate the new *assassin* character model. Furthermore, new unconventional characters could be designed by merging character attributes of different types together by joining different subgraphs – joining an *assassin* attributes subgraph with a *treasure hunter* subgraph to generate an *assassin treasure hunter* type character. The benefit of using a graph structure to represent characters, designers could traverse the graphs to create new unconventional characters. Apart from game designers, games can take the advantage of semantically-enriched character models and game character knowledge graphs to generate new unconventional non-playable characters on-the-fly which the user could interact with during the game. These new non-playable characters could also be generated based on the user’s experience and skills during the game – tougher characters would be generated for more skilled users whilst more lenient characters will be generated for less skilled users.

With the novel approach proposed in this paper we envisage not only the generation of characters in digital games autonomously but also the creation of games that are perceived as being unconventional and unexpected, yet engaging and playable. As future directions, we will expand...
this work to further improve the information extraction methodology to continue build a video game character knowledge graph. Moreover, we will improve our information extraction methodology to extract other video game features to collect more game content other than video game characters, such as levels and gameplay. We will also develop an automatic game character generator that will generate new unconventional game characters from the semantic annotated game character knowledge graph, and other generators to generate new unconventional video game content such as levels.

Acknowledgements

The research work disclosed in this publication is partially funded by the REACH HIGH Scholars Programme --- Post-Doctoral Grants. The grant is part-financed by the European Union, Operational Programme II --- Cohesion Policy 2014-2020 Investing in human capital to create more opportunities and promote the wellbeing of society --- European Social Fund [ESF.03.009].

References

[1] Assassin’s Creed Series. http: //assassinscreed.ubi.com.
[2] Baldur’s Gate Series. https: //www.baldursgate.com.
[3] DBpedia Ontology. http: //dbpedia.org/ontology/.
[4] Fandom. http: //fandom.wikia.com/games.
[5] Giant Bomb. http: //www.giantbomb.com/characters/.
[6] Neverwinter Nights. http: //www.bioware.com/en/games.
[7] Steam Discussions. http: //steamcommunity.com/discussions/.
[8] The Elder Scrolls Series. https: //elderscrolls.bethesda.net/.
[9] Tomb Raider Series. https: //www.tombraider.com.
[10] Uncharted Series. http: //www.unchartedthegame.com.
[11] G. A. Barros, A. Liapis, and J. Togelius. Playing with Data: Procedural Generation of Adventures from Open Data. In International Joint Conference of DLGRA and FGd, DGRA-FDG’16, 2016.
[12] G. A. Barros, A. Liapis, and J. Togelius. Who Killed Justin Bieber? Murder Mystery Generation from Open Data. In International Conference on Computational Creativity, ICCC’16, 2016.
[13] T. Berners-Lee. Semantic Web Road Map, September 1998.
[14] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. Scientific America, 284: 34–43, 2001.
[15] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee. In Linked Data on the Web (LDOW2008), 2008.
[16] K. Bontcheva, V. Tablan, D. Maynard, and H. Cunningham. Evolving GATE to Meet New Challenges in Language Engineering. Natural Language Engineering, 10 (3/4): 349–373, 2004.
[17] T. Brger, P. Hofmair, and G. Kienast. The salero virtual character ontology. In Proceedings of the First Workshop on Semantic 3D Media, 2008.
[18] D. Brickley and L. Miller. FOAF Vocabulary. http: //xmlns.com/foaf/0.1/.
[19] C. Browne and F. Maire. Evolutionary Game Design. IEEE Transactions on Computational Intelligence and AI in Games, 2(1):1–16, 2010.
[20] L. Carr, W. Hall, S. Bechofer, and C. Goble. Conceptual linking: ontology-based open hypermedia. In Proceedings of the 10th international conference on World Wide Web, pages 334–342. ACM, 2001.
[21] J. T. C. Chan and W. Y. F. Yuen. Digital Game Ontology: Semantic Web Approach on Enhancing Game Studies. In International Conference on Computer-Aided Industrial Design and Conceptual Design, CAID/CD 2008, 2008.
[22] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. A framework and graphical development environment for robust nlp tools and applications. In ACL, pages 168–175, 2002.
[23] B. O. Durić and M. Konecki. Specific owl-based rpg ontology. In Central European Conference on Information and Intelligent Systems, 2015.
[24] B. Farias Lsiao, C. Burle, and N. Calegari. W3C. Data on the Web Best Practices. 19 May 2016. W3C Working Draft. http: //www.w3.org/TR/dwbp/.
[25] T. Heath and C. Bizer. Linked Data: Evolving the Web into a Global Data Space. Morgan and Claypool, 2011.
[26] J. Kahan, M.-R. Koivunen, E. Prud’Hommeaux, and R. R. Swick. Annoteca: an open rdf infrastructure for shared web annotations. Computer Networks, 39 (3): 589–608, 2002.
[27] J. Kessing, T. Tutenel, and R. Bidarra. Designing semantic game worlds. In Workshop on Procedural Content Generation in Games, PGC’12. ACM, 2012.
[28] A. Liapis, G. N. Yannakakis, and J. Togelius. Computational game creativity. In Fifth International Conference on Computational Creativity, ICCC’14, 2014.
[29] R. Lopes and R. Bidarra. A semantic generation framework for enabling adaptive game worlds. In International Conference on Advances in Computer Entertainment Technology. ACM, 2011.
[30] C. D. Manning, H. Schu’tz et al. Foundations of statistical natural language processing, volume 999. MIT Press, 1999.
[31] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60, 2014.
[32] D. Maynard, V. Tablan, C. Ursu, H. Cunningham, and Y. Wilks. Named entity recognition from diverse text types. In Recent Advances in Natural Language Processing 2001 Conference, pages 257–274, 2001.
[33] 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.
[34] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, and A. Kirilov. Kim—a semantic platform for information extraction and retrieval. Natural language engineering, 10 (3-4): 375–392, 2004.

[35] O. R. Rocha and C. Faron-Zucker. Ludo: An Ontology to Create Linked Data Driven Serious Games. In ISWC 2015 - Workshop on LINKed EDucation, LINKED’15, 2015.

[36] O. Sacco. Game Character Ontology (GCO). http://autosemanticgame.eu/ontologies/gco#.

[37] 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.

[38] O. Sacco, A. Liapis, and G. N. Yannakakis. A Holistic Approach for Semantic-Based Game Generation. In Proceedings of the IEEE Computational Intelligence and Games Conference, CIG16, 2016.

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

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

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

[42] R. Warren and A. Dean-Hall. Appearance Ontology. http://rdf.muninn-project.org/.

[43] 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.

[44] 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.

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