Refining the Relationships among Historical Figures by Implementing Inference Rules in SWRL

Nurul Fajrin Ariyani, Madis Saralita, Sarwosri, and Riyanto Sarno
Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: nurulfajrin@if.its.ac.id, madis.saralita@gmail.com

Abstract. The biography of historical figures is often fascinating to be known. Everything about their character, work, invention, and personal life sometimes are presented in their biography. The social and family relationships among historical figures also put into concern, especially for political figures, heroes, kings or persons who have ever been ruled a monarchy in their past. Some biographies can be found in Wikipedia as articles. Most of the social and family relationship contents of these figures are not completely depicted due to a various article’s contributors and sources. Fortunately, the missing relatives of a person might reside in the other figures’ biography in different pages. Each Wikipedia article has metadata which represents its essential information of content. By processing the metadata obtained from DBpedia and composing the inferencing rules (in the form of ontology) to identify the relationships content, it can generate several new inferred facts that complement the existing relationships. This work proposes a methodology for finding missing relationships among historical figures using inference rules in an ontology. As a result, our method can present new facts about the relationships that absent in the existing Wikipedia articles.

1. Introduction
A historical figure is a person whose name is remembered for his or her action, influence, or involvement in particular historical events in the past. The biography of historical figures have various characters of information depends on their types. Some kings have had a few noble titles due to their rule as a king, queen, or emperor of a monarchy. However, some of other historical figures have labeled as heroes because of their heroic acts in the past events which affected the history nowadays. A historical figure is an evidence of a notable event that happened in particular time. Each figure has his or her own story and personal life which had involved the other figures based on his or her social and family relationships. The relationships among figures could depict the kind of historical event that had involved them in the past.

By knowing the family relationships among kings, we will able to build a complete genealogy of them which leads to the understanding of throne inheritance and the origin of the kingdoms. While by knowing the social relationships, we will know the alliance of a historical figure, what political parties he or she was involved and whether some figures were lived and cooperate in the same period of time. Both subjects are parts of history learning materials in schools. Thus, adding relationships among historical figures can give contributions to enrich knowledge learning materials for students.

The biography of historical figures often presents their social and family relationships with the other figures. A free online dictionary such as Wikipedia has preserved a large number of historical
figure’s biographies in the form of article pages. The relationships information that depicted in Wikipedia biographies may appear incomplete due to the different article’s contributors or sources. Moreover, the name of historical figure tends to have multiple aliases which can burden the process of manually finding the missing relationships among them. Each Wikipedia article has metadata in the form of RDF format that stored in DBpedia. Fortunately, these missing relationships can be inferred from the other related pages by processing articles’ metadata based on predefined inference rules that created in SWRL.

However, the inference rules reasoning new facts about relationships use specific terms that defined in an ontology. Thus, to be able to implement the rules, the original facts (existing Wikipedia metadata) need to be modeled using an ontology. An ontology defines classes, properties, and instances of a particular domain. Actor is one of the domains in cultural heritage ontology. The scope of actor involves a person, group, and organization. Historical figures including kings and heroes are categorized in person agent.

An ontology was built for specific purposes, and only valid within its domain. To create an ontology, we shall modify or integrate several existing ontologies rather than built it from scratch [1]. Several studies about building metadata of cultural heritage, including person were conducted by adopting CIDOC-CRM ontology which focuses on the relation between event, temporal, and spatial [2,3]. However, to find the relationships among historical figures, we need to combine several ontologies that related to biography and friendship relations. Our proposed ontology was developed in Ontology Web Language (OWL) using Protégé as the ontology code editor tool.

2. Related Works

In the previous works, a Biography Light Ontology (BLO) was built by composing several well-known ontologies such as ABC Ontology, Linking Open Description of Events Ontology (LODE), Friend of a Friend (FOAF), BIO Vocabulary, Relationship Vocabulary, CIDOC-CRM and Bibliographic Ontology (Bibo). This ontology’s structure is more dynamic to represent biographical events rather than other models [4]. However, it lacks some specific terms regarding to describe a group, family, and organization representation. Its interoperability between the existing specifications and relevant standards for biography encoding still need improvement.

The other ontology, LODE, provides comprehensive evaluations of event ontology and Interlingua model to enhance the interoperability among the existing standards. LODE is an event-centric ontology which relates some instances based on their involvement in the same single event [5]. But in fact, the socio family relationships among historical figures that represented in DBpedia biography were not complemented with information about the figure’s participation in an event, even though some specific events may explicitly appear in the metadata.

![Figure 1. The Process Diagram](image-url)
FOAF is a project intended for connecting between a person and his or her information via the web. FOAF is more suitable to depict a network between people in the modern life which has been influenced by social media technology [6]. However, FOAF is still used practically as a fundamental ontology and often combined with other well-known ontologies such as Bibo and DBpedia [7].

One of the other commonly used biography ontology named BIO. BIO document consists of a vocabulary to model biography information about a person and his or her background [8]. BIO provides a class for event and properties to relate a person to event and event to its place and time [9]. BIO defines several relationships such as Father, Mother, Child, Birth, Death, and so forth. But it is not sufficient to describe other family relationships outside the primary family member. To complement the relationship properties in the family domain, we used FamilyTree that was derived from University of Manchester’s portal1. FamilyTree is one of the complex and comprehensive ontologies in the domain of family. The main purpose of this ontology is to generate new facts by minimizing relationships (property usage) and optimizing inference (rules). However, to enhance the inference capability, we added several SWRL rules to support the axioms that had been decided in the Family Tree. SWRL rules have flexibility in representing rule facts in common condition which including text and math operations also can be implemented in many cases other than genealogy [10].

Several works related to modeling the genealogy or family relationships using SWR rules to enriching the knowledge have done for many purposes, mostly for medical issues. Paper [11] employed GenOnto to structure the relatives of a patient that tend to suffer from Li-Fraumeni Syndrome. Its GenOnto covers until the third degree of relatives, i.e., cousins. An application for helping elderly to remember their family, called CAPTAIN MEMO, also implemented an ontology and inference rules to represent family relationships [12]. The genealogy domain is suitable for defining the real world semantics as depicted in [13]. In this paper, the authors implemented several axioms in their reference domain to determine the memberships of its classes. Based on this works, we can conclude that inference rules contribute in enriching the family relationships by producing inferred facts from existing axioms.

The ontology proposed in this work merged FamilyTree, FOAF, and BIO by performing the classes and properties combination. The merged ontology was developed by adding several new classes and properties to facilitate terms that could not be accommodated by the three original ontologies.

3. Methodology
The merger and development of the ontology were aimed to produce an ontology that can define the relationship type among person. Figure 1 shows the process diagram implemented in this work. Each process is described in the next following paragraph.

3.1. Data Analysis
Data analysis proceeded by taking several metadata of historical figures from DBpedia. To build a complete ontology that covers all things of the domain needs high cost due to the massiveness of data, time processing, and the inevitably changing facts over times [14]. Thus, some historical figures (person) that observed as data are people who considered of having a lot of relationships with others. Each person has biography related to place (e.g., BirthPlace and DeathPlace), time (e.g., BirthDate and DeathDate), and event (e.g., Perang_Aceh, Proklamasi, etc).

At first, we tried to create a metadata manually from Wikipedia pages. The label and term produced in this process did not have any standardization which would be difficult to merge in the later phase. Next, we utilized WikiTaxi to catch the XML format of Wikipedia article. We found that this XML also could not be processed efficiently since its taggings were not structured in order. Due to this reasons, we decided to get the metadata from DBpedia comfortably.

---

1 http://www.co-ode.org/roberts/family-tree.owl
DBpedia provides the structure and content extraction (metadata) of Wikipedia articles. In fact, the metadata is composed of triples which describe any essential context on a web page as written below:

\[(\text{Subject}, \text{Predicate}, \text{Object})\]

The **Subject** represents any kinds of resources that exist in a world wide web and can be identified by its URI (Universal Resource Identifier). A **Predicate** links between subject and object in a particular context, but only limited to terms that listed in the vocabulary (ontology). It may have roles as an **object property** or a **data property**. A predicate that links between two URIs called object property. Otherwise, a predicate that links a URI with a literal value named data property. For example, there is a fact mentioned that Cut Nyak Dhien is the spouse of Teuku Umar. In DBpedia, this fact will be transformed into a triple as below.

\[(\text{CutNyakDhien}, \text{dbpedia}\text{-owl:spouse}, \text{TeukuUmar})\]

### Table 1. The List of Imported Classes and Properties

| **BIO**   | Classes                              | Object Properties                  |
|-----------|--------------------------------------|------------------------------------|
| **Classes** | Agent, Event, Group Event, Marriage, Individual Event, Birth, Burial, Death, Graduation, Person, Wedding, Event Generic | differentFrom, Agent, Partner, Principal |
| **FamilyTree** | Man, Woman, Sex, Male, Female | hasChild, hasSon, hasDaughter, hasParent, hasFather, hasMother, hasHusband, hasWife, isChildOf, isSonOf, isDaughterOf, isParentOf, isFatherOf, isMotherOf, inLawOf, isSpouseOf, isHusbandOf, isWifeOf |
| **Data Properties** | hasName, hasFirstName, hasFamilyName, knownAs, alsoKnownAs, formerlyKnownAs, hasEventYear, hasBirthday, hasDeathYear |

#### 3.2. Ontology Developing

In the ontology design, we merged three different kind ontologies which are FamilyTree, FOAF, and BIO. The merged ontology, which named **iHero**, later has been adjusted to preserve some information of the biography’s metadata. The terms which came from two different ontologies but have a similar domain could be integrated as a contextual information. However, if there is still any information that could not be addressed by the integrated ontology, we may add new terms to refine it [15].

#### 3.3. Importing Classes and Properties

The classes and properties of iHero were carefully chosen from three original ontologies by considering the context of each term that usually described in the annotation. To ensure that the related terms have the same meaning, we also utilized **SUMO**, a credible upper ontology that widely used in the ontology merging [16]. The list of classes and properties with its originality is shown in Table 1.

#### 3.4. Equivalent and Invers

The equivalent classes are two or more classes which have exactly same meaning or context. This equivalent attribute was derived from the original ontology. In iHero ontology, **Marriage** class is equivalent to **WeddingEvent Generic** class. Meanwhile, for its properties, **hasChild** property is equivalent to **isParentOf** property, and **hasParent** is equivalent to **isChildOf**. For example, if there is a fact about Cut Gambang **hasParent** Teuku Umar, it also means that Cut Gambang **isChildOf** Teuku Umar.
family genealogy rules in SWRL. Each of the family relationship rules was derived from several
is no automatic way to generate the inference rules; hence they need to be built by hand by code the
process.

The purposes of designed rules are to find relationships between one person to others. Rules were built
For example, the rule
We created 21 rules to find social family relations, as mentioned in Table 2. Please note that there

| No  | Property        | Rule                                                                 |
|-----|-----------------|----------------------------------------------------------------------|
| R1  | isSpouseOf      | Marriage(m) ∧ Female(sf) ∧ Male(sm) ∧ Partner(m, sf) ∧ Partner(s, sm) → isSpouseOf(sm, sf) |
| R2  | has Husband     | Marriage(m) ∧ Female(sf) ∧ Male(sm) ∧ Partner(m, sf) ∧ Partner(m, sm) → has Husband(sm, sf) |
| R3  | has Wife        | Marriage(m) ∧ Female(sf) ∧ Male(sm) ∧ Partner(m, sm) ∧                     |
|     |                 | has Wife(sm, sf)                                                         |
| R4  | has Father      | Male(m) ∧ Person(p) ∧ has Parent(p, m) → has Father(p, m)               |
| R5  | has Mother      | Female(f) ∧ Person(p) ∧ has Parent(p, f) → has Mother(p, f)              |
| R6  | has Child       | Female(f) ∧ Male(m) ∧ Person(p) ∧ has Parent(p, f) ∧ has Child(f, p) ∧ has Child(m, p) |
| R7  | has Son         | Male(m) ∧ Person(p) ∧ has child(p, m) → has Son(p, m)                   |
| R8  | has Daughter    | Female(f) ∧ Person(p) ∧ has Child(p, f) → has Daughter(p, f)            |
| R9  | has Sibling     | Female(f) ∧ Male(m) ∧ Person(p) ∧ has Parent(p, f) ∧ has Sibling(p, f)   |
| R10 | has Brother     | Male(m) ∧ Person(p) ∧ has Sibling(p, m) → has Brother(p, m)             |
| R11 | has Sister      | Female(f) ∧ Person(p) ∧ has Sibling(p, f) → has Sister(p, f)            |
| R12 | has Step Father | Person(c) ∧ Person(p1) ∧ Person(p2) ∧ Person(p3) ∧ has Husband(p1, p2) ∧ |
|     |                 | has Husband(p1, p3) ∧ has Parent(p1, p1) ∧ has Parent(p2, p2) ∧ Diff   |
|     |                 | rentFrom(p2, p3) → has StepFather(c, p3)                                |
| R13 | has Step Mother | Male(m) ∧ Person(p) ∧ has Step Parent(p, m) → has Step Father(p, m)     |
| R14 | has Step Son    | Female(f) ∧ Person(p) ∧ has Step Child(p, m) → has Step Mother(p, f)   |
| R15 | has Step Daughter | Female(f) ∧ Person(p) ∧ has Step Child(p, f) → has Step Daughter(p, f) |
| R16 | has Step Sibling | Person(c) ∧ Person(c1) ∧ Person(p) ∧ has Child(p, c2) ∧ has Step Parent |
|     |                 | (c1, p) → has Step Sibling(c1, c2)                                        |
| R18 | has Step Brother | Male(m) ∧ Person(p) ∧ has Step Sibling(p, m) → has Step Brother(p, m)  |
| R19 | has Step Sister | Female(f) ∧ Person(p) ∧ has Step Sibling(p, f) → has Step Sister(p, f) |
| R20 | is Ally Of      | Event(w) ∧ Group(g) ∧ Person(p1) ∧ Person(p2) ∧ has Alleloegy(p1, g) ∧  |
|     |                 | has Alleloegy(p2, g) ∧ has Commander(w, p1) ∧ has Commander(w, p2) ∧   |
|     |                 | DifferentFrom(p1, p2) → is Ally Of(p1, p2)                              |
| R21 | is Enemy Of     | Event(w) ∧ Group(g1) ∧ Group(g2) ∧ Person(p1) ∧ Person(p2) ∧ has Alleloegy(p1, g1) ∧  |
|     |                 | has Alleloegy(p2, g2) ∧ has Commander(w, p1) ∧ has Commander(w, p2) ∧   |
|     |                 | DifferentFrom(p1, p2) → is Enemy Of(p1, p2)                            |

Same as equivalent attribute, an inverse one also can be attached to classes or properties. However, it has an opposite meaning with the equivalent attribute. For example, hasChild property is an inverse of hasParent property. Thus, if there is fact about Cut Nyak Dhien hasParent Cut Gambang, it also means that Cut Gambang hasParent Cut Nyak Dhien.

3.5. Build SWRL Rules
The purposes of designed rules are to find relationships between one person to others. Rules were built in SWRL (Semantic Web Rule Language) by composing a pair of antecedent and consequent [17]. If the value of antecedent is true, then its consequence’s value will also be true. The rules were expected to be able to create new facts related to a figure that is still absent in DBpedia by doing inferencing process.

We created 21 rules to find social family relations, as mentioned in Table 2. Please note that there is no automatic way to generate the inference rules; hence they need to be built by hand by code the family genealogy rules in SWRL. Each of the family relationship rules was derived from several family history term in [18].

For example, the rule R21 will be executed by utilizing class Event, Group, and Person. A single
Event(w) which is probably an event of war, two groups Group(g1) and Group(g2) which each represents a nation, and two people Person(p1) and Person(p2) are involved in this rule as mentioned
in Table 2. Person\( (p1) \) has devoted to nation Group\( (g1) \), and otherwise Person\( (p2) \) has devoted to nation Group\( (g2) \). Both people have participated as the commander in the same war Event\( (w) \). However, Person\( (p1) \) and Person\( (p2) \) are different individuals. Also, they both have devoted to different Group. If all of these conditions are met in the existing facts, then it will produce a new information that Person\( (p1) \) is the enemy of Person\( (p2) \).

4. Evaluation

We used 20 meta data of historical figure’s biographies as our experimental data regarding to their completeness of properties in DBpedia. These 20 figures are: Soekarno, Fatmawati, Mohammad Hatta, Hasyim Asyari, Mohammad Natsir, Cut Nyak Dhien, Teuku Umar, Cut Nyak Meutia, Teuku Nyak Arif, Pierre Tandeant, Abdul Haris Nasution, Ahmad Yani, Ahmad Dahlan, Nyai Ahmad Dahlan, Tuanku Imam Bonjol, Wahid Hasyim, Mas Mansoer, Diponegoro, Sudirman, dan Slamet Rijadi. We have evaluated each of the 21 rules built in SWRL to ensure the correctness of its inferred facts derived from existing data. We applied Pellet Reasoner as the inferencing or reasoning engine and utilized Protégé to for editing and querying individuals. The complete results of evaluation are shown in Table 5.

The second column in Table 5 depicts one sample of the inferred facts resulted from reasoning, while in the third column is the correctness status of inferred facts that thoroughly compared with actual data from other sources. Several figures in second column are followed by their Wikipedia’s article ID for clearness purposes. However, there are also other some figures that are not followed by IDs. It means that these figures were inferred from the reasoning process (not directly derived from a Wikipedia article).

For example, to evaluate the rule of isEnemyOf we compared between existing facts and inferred facts of Sudirman as shown in Table 3 and Figure 2. One of the inferred facts mentioned that Sudirman isEnemyOf William Franken as result of the rule in R21. William Franken hasAllegiance with Netherlands as this fact is explicitly written in Wiliam Franken’s biography. Both Sudirman and William Franken were commander of Indonesian National Revolution event, though, they were not joined in the same allegiance. Due to this matters, the reasoner concluded that Sudirman isEnemyOf William Franken.

For example of family relationships rule, we put Cut Nyak Dhien biography as a sample. Table 4 shows several existing facts of this figure (please note that we only cropped properties related to family relations for clearness purposes). In Figure 3, it appeared that Cut Nyak Dhien hasStepChild Teuku Raja Sulaiman. This inferred fact resulted from hasStepParent rule in R12 and the inverse properties between hasStepParent and hasStepChild. Teuku Raja Sulaiman is the child of Teuku Umar and Cut Meuligou, Teuku Umar’s wife from the previous marriage. Since Cut Meuligou and Cut Nyak Dhien are different individuals, the reasoner executed the rule and concluded that Teuku Raja Sulaiman hasStepParent Cut Nyak Dhien. Since hasStepChild is inverse of hasStepParent property, so it also produced that fact.

![Table 3. Existing Metadata of Sudirman](image1)

| Elements | Property Name | Value |
|----------|---------------|-------|
| hasAllegiance | G_Indonesia |
| isCommanderOf | E_Indonesian_National_Revolution |
| isCommanderOf | Agresi Militer Belanda’I |
| isCommanderOf | E_Battle_of_Ambawara |
| isInfluencedBy | E_Tau_Malaka |
| hasSubject | S_Indonesian_generals |

![Table 4. Existing Metadata of Cut Nyak Dhien](image2)

| Elements | Property Name | Value |
|----------|---------------|-------|
| hasSpouse | P_Teuku_Umar |
| hasChild | P_Cut_Gambang |
| isCommanderOf | E_Aceh_War |
| knownFor | G_National_Heroine |
As we see from Table 5, we may conclude that all of our inference rules were successfully implemented and could produce new inferred facts based on the rules. However, the isAllyOf and isEnemyOf properties need to be revised further since a figure might change his/her allegiance in the past. We consider employing the timeline sequence to clear the validity period of each inferred facts. Furthermore, the reasoning process seems to be not efficiently implemented since we need to maximize the heap size memory of Protégé to complete the whole process. By utilizing 2 GB of memory, the reasoning process was completed in less than a minute. However, the time needed to export all of the inferred axioms (so it will be able to be queried) was more than 20 minutes.

5. Conclusion
We have developed an ontology for finding relationships among historical figures by merging FOAF, BIO, and FamilyTree scheme and building SWRL rules to reason the missing relations. Rules’ functionalities could produce several new inferred facts based on the existing facts in the 20 data figures. Limited memory used during processing affects the complexity of reasoning process, though...
its need a lot more time to export the inferred axioms into physical files so they could be queried using Protégé’s SPARQL endpoint.

Regardless of the success of the proposed ontology, there is a lack of methods in the reasoning process. The reasoning processing time will grow exponentially to the complexity of defined rules in the ontology. Our future works aim at this challenge of efficiently reasoning metadata, so it will be appropriate to implement in the real web based application for searching relationships of historical figures.

Acknowledgments
This research funding was granted by Lembaga Penelitian dan Pengabdian Masyarakat, Institut Teknologi Sepuluh Nopember Surabaya as Penelitian Pemula Dana Lokal ITS 2017.

References
[1] Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing?. International journal of human-computer studies, 43(5-6), pp.907-928.
[2] Ariyani, N.F., Hanafi, M.B., Ginardi, H. and Saralita, M., 2016, October. Linking spatial information of cultural heritage metadata with geo linked open data. In Information Technology Systems and Innovation (ICITSI), 2016 International Conference on (pp. 1-5). IEEE.
[3] Ariyani, N.F. and Yuhana, U.L., 2015, November. Generating cultural heritage metadata as linked open data. In Information Technology Systems and Innovation (ICITSI), 2015 International Conference on (pp. 1-6). IEEE.
[4] Ramos, M.R., Biography Light Ontology: An Open Vocabulary For Encoding Biographic Texts.
[5] Shaw, R., Troney, R. and Hardman, L., 2009. LODE: Linking Open Descriptions of Events. ASWC, 9, pp.153167.
[6] Al-Mukhtar, M.M.A. and Al-Assafy, A.T.A., 2014. The implementation of foaf ontology for an academic social network. International Journal of Science, Engineering and Computer Technology, 4(1), p.10.
[7] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S. and Bizer, C., 2015. DBpedia: a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web, 6(2), pp.167-195.
[8] Davis and Galbraith, D., BIO: A vocabulary for biographical information. [Online]. Available: http://vocab.org/bio/. [Accessed 06 January 2016].
[9] Larson, R., Bringing lives to light: Biography in context (2010).
[10] Hartanto, H.A., Sarno, R. and Ariyani, N.F., 2016, August. Linked warning criterion on ontology-based key performance indicators. In Technology of Information and Communication (iSemantic), International Seminar on Application for (pp. 211-216). IEEE.
[11] Budaruiche, R.M., Wassermann, R., Patrao, D.F. and Achatz, M.I., 2015, November. Li-Fraumeni Ontology: A Case Study of an Ontology for Knowledge Discovery in a Cancer Domain. In Intelligent Systems (BRACIS), 2015 Brazilian Conference on (pp. 157-162). IEEE.
[12] Herradi, N., Hamdi, F., Mtais, E., Ghorbel, F. and Soukane, A., 2015, October. PersonLink: an ontology representing family relationships for the CAPTAIN MEMO memory prosthesis. In International Conference on Conceptual Modeling (pp. 3-13). Springer, Cham.
[13] De Carvalho, V.A., Almeida, J.P.A. and Guizardi, G., 2014, June. Using reference domain ontologies to define the real-world semantics of domain-specific languages. In International Conference on Advanced Information Systems Engineering (pp. 488-502). Springer, Cham.
[14] Wu, D. and Hkansson, A., 2014. A method of identifying ontology domain. Procedia Computer Science, 35, pp.504-513.
[15] Simperl, E., Sarasua, C., Ungrangsi, R. and Brger, T., 2011. Ontology metadata for ontology
reuse. *International Journal of Metadata, Semantics and Ontologies*, 6(2), pp.126-145.

[16] Niles, I. and Pease, A., 2001, October. Towards a standard upper ontology. In Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001 (pp. 2-9). ACM.

[17] Oconnor, M., Knublauch, H., Tu, S. and Musen, M., 2005. Writing rules for the semantic web using SWRL and Jess. *Protégé With Rules WS, Madrid*.

[18] Watson, J., 1996. *Ordering the family: Genealogy as autobiographical pedigree*. na.