TatWordNet: A Linguistic Linked Open Data-Integrated WordNet Resource for Tatar

Alexander Kirillovich ✉ ⓘ
Kazan Federal University, Kazan, Russia
Higher School of Economics, Moscow, Russia

Marat Shaekhov ✉ ⓘ
Kazan Federal University, Kazan, Russia

Alfiya Galieva ✉
Kazan Federal University, Kazan, Russia

Olga Nevzorova ✉ ⓘ
Kazan Federal University, Kazan, Russia
Higher School of Economics, Moscow, Russia

Dmitry Ilvovsky ✉ ⓘ
Kazan Federal University, Kazan, Russia
Higher School of Economics, Moscow, Russia

Natalia Loukachevitch ✉ ⓘ
Moscow State University, Moscow, Russia
Kazan Federal University, Kazan, Russia
Higher School of Economics, Moscow, Russia

---- **Abstract** ----

We present the first release of TatWordNet (`http://wordnet.tatar`), a wordnet resource for Tatar. TatWordNet has been constructed by the combination of the expand and the merge approaches. The synsets of TatWordNet have been compiled by: (i) the automatic conversion of concepts of TatThes, a socio-political Tatar; (ii) semi-automatic translation of synsets of RuWordNet, a wordnet resource for Russian with the followed manual verification and correction; (iii) manual translation of base RuWordNet synsets; (iv) and manual translation of the all hypernyms of the previously translated RuWordNet synsets. The currents version of TatWordNet contains 18,583 synsets, 36,540 lexical entries and 49,525 senses. The resource has been published to the Linguistic Linked Open Data cloud and interlinked with the Global WordNet Grid.

1 Introduction

The Princeton WordNet thesaurus (PWN) [9, 11] is one of the most important language resources for linguistic studies and natural language processing. PWN is a large-scale lexical knowledge base for English, organized as a semantic network of synsets. A synset is a set of words with the same part-of-speech that can be interchanged in several contexts. Synsets are interlinked by semantic relations, such as hyponymy (between specific and more general concepts), meronymy (between parts and wholes), antonymy (between opposite concepts) and other.

Inspired by success of PWN, many projects have been initiated to develop wordnets for other languages across the globe. Nowadays wordnet-like resources are developed for nearly 80 languages, but Tatar language is not among them. In this paper, we fill the gap and present the first release of TatWordNet (`http://wordnet.tatar`), a wordnet resource for Tatar.

2   Related works

At present time, there are various wordnets for some Turkic languages.

Two Turkish wordnet projects have been developed for the Turkish language. The first one [5, 2] has been created at Sabancı University as part of the BalkaNet project [19]. The BalkaNet project was built on the basis of a combination of expand and merge approaches. All wordnets contain many synonyms for Balkan common topics, as well as synsets typical for each of the BalkaNet languages. The size of Turkish Wordnet is about 15,000 synsets.

Another Turkish wordnet is the KeNet [1, 7, 8]. This wordnet was built on the basis of modern Turkish dictionaries. To build this resource, a bottom-up approach was used. Based on dictionaries, words were selected and then manually grouped into synsets. The relationships between words have been automatically extracted from dictionary definitions and then the latter have been fixed between synsets. The size of this resource is about 113,000 synsets.

Unfortunately, lack of large Turkish-Tatar dictionaries (as well as English-Tatar ones) makes it impossible to translate Turkish resources into the Tatar language. In this respect, the Tatar language can be attributed to low-resource languages.

The Extended Open Multilingual Wordnet [3] resource is built from Open Multilingual Wordnet by replenishing the WordNet data automatically extracted from the Wiktionary and Unicode Common Locale Data Repository (CLDR). The resource contains wordnets for 150 languages, including several Turkic: Azerbaijani, Kazakh, Kirghiz, Tatar, Turkmen, Turkish, and Uzbek. The Tatar wordnet contains a total of 550 concepts, which covers 5% of the PWN core concepts.

The BabelNet [18] resource contains a common network of concepts that have text inputs in many languages. The BabelNet contains 90,821 Tatar text entries that refer to 63,989 concepts. However, due to the fact that this resource was built automatically, it has quality issues. Thus, the development of a quality Tatar wordnet with an emphasis on the specific features of the Tatar language based on the existing lexical resources is very relevant.

3   TatWordNet construction

There are two main approaches for construction of wordnets for new languages: expand and merge [20]. The expand approach is to take the semantic network of PWN and translate its synsets into the target language, adding additional synsets when needed. The merge approach is to develop a semantic network in the target language from scratch and then link it to PWN.

Since the merge approach is very labor-intensive and time consuming, the expand approach seems more appropriate for under-resources languages such as Tatar. However, in development of TatWordNet, the expand approach can't be directly applied either, due to the lack of large English-Tatar dictionaries, necessary for translation of PWN to Tatar. At the same time, there are several relatively large and high-quality Russian-Tatar dictionaries, so Russian thesauri can be used as the source resources instead of PWN.
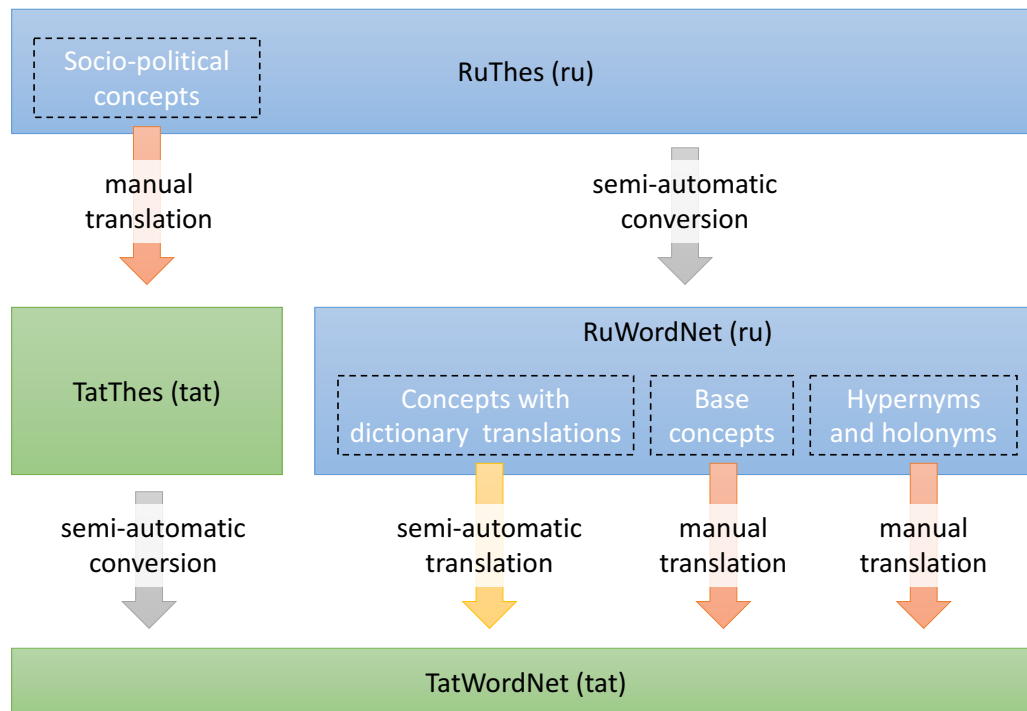
Figure 1 TatWordNet development.

With this consideration in mind we constructed TatWordNet on the base of three source resources, developed by us: RuThes, RuWordNet and TatThes. In this section we describe the resources, that were used to produce TatWordNet and then the construction process itself (Fig. 1).

RuThes

RuThes [16, 15, 13] is a thesaurus and a linguistic ontology for Russian. It is organized as a network of concepts, that are considered as language-independent "units of thought".

The concepts are language-independent in the sense that their identities and distinctions from each other don't depend on the terms that express them. At the same time, the network of concepts is linguistically motivated, i.e., it contains mostly those concepts, that are denoted by actual language expressions.

A concept is characterized by a unique name, and optionally by a gloss. Every concept is associated with lexical entries, by which this concept is referred to. The lexical entries associated with the same concept are called ontological synonyms. Ontological synonyms can comprise:

- words belonging to different parts of speech;
- language expressions relating to different linguistic styles and genres;
- idioms and free multiword expressions.

For example, the list of the lexical entries for the concept *Surgical operation* includes:

- the noun "операция" ("operation");
- the verbs "оперировать" ("to operate", imperfective) and "прооперировать" ("to operate", perfective);

- the adjectives "операционный" ("operative") and "хирургический" ("surgical");
- the noun phrases "хирургическая операция" ("surgical operation"), "хирургическое лечение" ("surgical treatment") and "оперативное вмешательство" ("operative intervention");
- the verb phrase "оперировать пациента" ("to operate on a patient");
- and the idiom "лечь под нож" ("to go under the knife").

Just like in WordNet, an ambiguous lexical entry is assigned to several concepts. For example, the word "коса" is assigned to three different concepts: *Tongue of land, Braid of hair* and *Scythe.*

RuThes defines four main relations between concepts:

1) The taxonomic relation, that is a union of the traditional ontological class-subclass (isA) and instanceOf relations. For example, this relation holds between the *Moscow* and the *City* concepts as well as between the *City* and *Settlement* concepts. The RuThes taxonomic relation is analogous to the hyponym-hypernym relation of WordNet.

2) The part-whole relation. In RuThes, this relation is interpreted fairly broadly, and applies to entities of many different types, including:

- physical objects (*Car engine – Car*), regions (*Europe – Eurasia*), substances, sets (*Battalion – Company*);
- processes (*Public prosecution – Judicial trial*);
- an attribute and its bearer (*Displacement – Ship*);
- a role or a participant of a situation and the situation (Teacher – Education);
- entities and situations in the encompassing sphere of activity (*Industrial plant – Industry, Tennis racket – Tennis, Tennis player – Tennis*).

At the same time the RuThes part-whole relation has a very important restriction: a concept-part should be related to its whole during normal existence of its instances. For example, the Tree concept is not described as part of the Forest concept, because trees can grow in many places, not only in forests. This makes it possible to use the transitivity of the part-whole relations with greater reliability.

The RuThes part-whole relation is similar to the WordNet meronym-holonym relation, and being applied to spheres of activity concepts it also can be used to model the WordNet domain relation.

3) The directed association relation, that expresses the relation of the external ontological dependence between two concepts. The association relation is established between two concepts C_1 and C_2 when C_1 ontologically depends on C_2 and C_1 is not a part of C_2. For example, this relation holds between the *Auto racing* and *Car* concepts.

4) The undirected symmetric association relation.

The RuThes relations have formal-ontological nature, allowing them to be subjected to the following formal inference rules: (1) transitivity of the part-whole relations; (2) inheritance of the part-whole relationships to subclasses; (3) inheritance of association relationships to parts and subclasses.

RuThes has considerable similarities with WordNet: both resources are composed of concepts/synsets, that are organized into a network by predefined set of conceptual relations, and associated with semantically related lexical entries.

At the same time, there are several differences between RuThes and WordNet, the most important of which is that a RuThes concept can be associated with lexical entries, belonging to different part of speech. Due to these differences, RuThes is not fully compatible with some WordNet-oriented NLP applications.

In order to obtain a Russian resource fully compatible with WordNet standards, the RuThes developers transformed RuThes to RuWordNet, a WordNet-like resource for Russian.

RuWordNet

RuWordNet (RWN) [14, 17] is a Russian wordnet, semi-automatically generated from the RuThes thesaurus.

To create RuWordNet, the single conceptual network of RuThes was transformed to synsets' subnets for each part of speech, which were then enriched by additional wordnet-specific relations.

This transformation was carried out by the following steps:

1) At the beginning, every RuThes concept was divided into part-of-speech-related synsets, in such a way that each synset contains all the lexical entries of the source concept which belong to the corresponding part-of-speech (i.e. a noun synset contains all the noun lexical entries of the source concept, a verb synset contains all the verbs, and the same for an adjective synset).

Fig. 1 represents examples of dividing a RuThes concept into part-of-speech-related RWN synsets. The *Coronation* concept has noun, verb and adjective lexical entries, and so it was divided into three part-of-speech-related synsets:

- the noun synset {"коронация" ("coronation")};
- the verb synset {"короновать" ("to crown"), "возвести на престол" ("to enthrone")};
- and the adjective synset {"коронационный" ("coronational")}.

At the same time, the Induction concept has only noun and verb lexical entries, and so it was divided into two synsets.

The synsets, obtained from the same source concept, were linked to each other with the relation of part-of-speech synonymy.
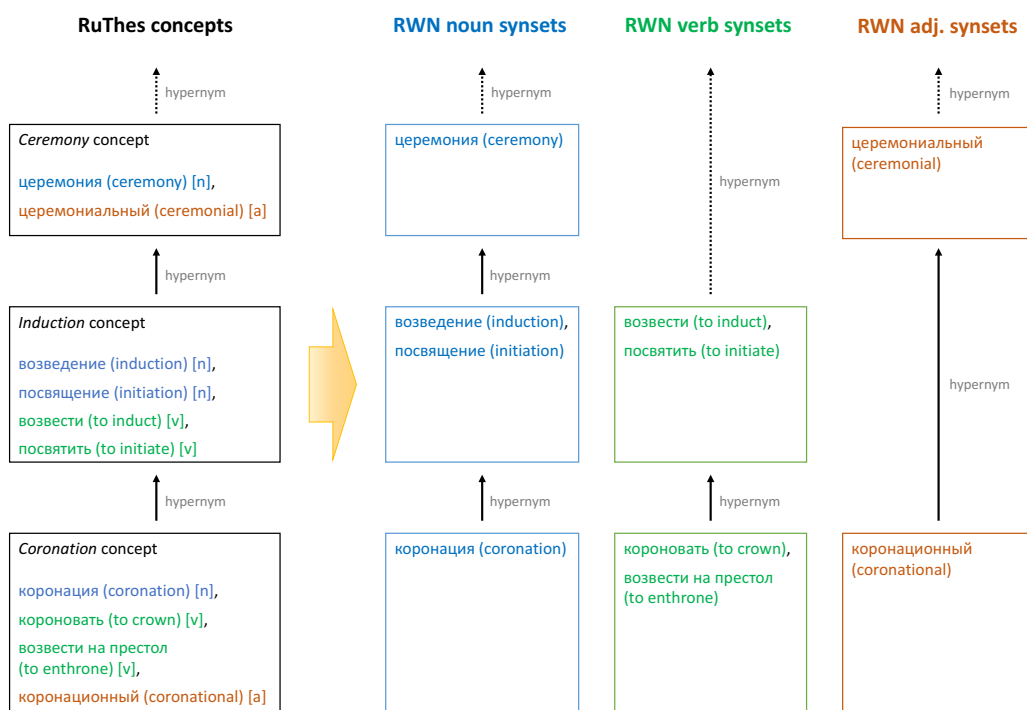
2) Then, the hyponym-hypernym and the meronym-holonym relations between RuThes concepts were reproduced for the corresponding RWN synsets of the same part-of-speech.

When two RuThes concepts, connected by the hyponym-hypernym relation, have corresponding RWN synsets of the same part-of-speech, the relation is established between these synsets.

It is, however, very common that in a pair of hyponym and hypernym concepts, one of the concepts does have a corresponding synset of a particular part-of-speech, but another concept doesn't. For example, the adjective synset {"коронационный" ("coronational")} is associated with the *Coronation* concept, but there is no adjective synset, associated with the *Coronation*'s hypernym, the *Induction* concept.

In such cases, hyponym-hypernym relation is established between two RWN synsets of the particular part-of-speech, if their source RuThes concepts are connected indirectly by the hyponym-hypernym path (and the intermediate concept in the path don't have themselves corresponding synsets of the relevant part of speech). For example, the relation is established between the aforementioned adjective synset {"коронационный" ("coronational")} and the adjective synset {"церемониальный" ("ceremonial")}, because their source RuThes concepts, *Coronation* and *Ceremony* respectively, are connected indirectly by a hyponym-hypernym path via the intermediate *Induction* concept (see Fig. 2).

In accordance with the WordNet standards, the hyponym-hypernym relation was additionally subdivided to the proper hyponym-hypernym and the instance hyponym-hypernym relations. In the current version of RuWordNet, the instance hyponym-hypernym relation was established for geographical objects.

Figure 2 Transformation of the RuThes conceptual network into three RWN subnets of synsets (fragment).

The meronym-holonym relation was reproduced for the RWN synsets in the same way as the hyponym-hypernym one. After that, it was semi-automatically corrected according to the WordNet standards.

3) Finally, the established relations were semi-automatically supplemented by several wordnet-specific relations, including the antonymy relation, the relations of causation and entailment, the domain relation, the relations of word derivation and the relations between phrases and their components. This process is not relevant to the TatWordNet development, and so we will not describe it in this review.

4) Additionally, RuWordNet was linked to Princeton WordNet via the Global WordNet inter-lingual-index (ILI).

TatThes

TatThes [10] is a socio-political thesaurus for Tatar, developed on the basis of the conceptual network of the RuThes thesaurus.

TatThes can be described as a kind of satellite resource for RuThes: it doesn't define its own conceptual network, but heavily reuses the conceptual network of RuThes, extending it by new Tatar-specific concepts and supplementing the existing RuThes concepts by Tatar lexical entries. (It should be noted, that the reused RuThes concepts are defined only once in the RuThes itself, and TatThes only refers to them in accordance to the Linked Open Data principles).

Table 1 TatWordNet construction statistics.

Step	# of synsets
TatThes automatic conversion	4,422
RWN semi-automatic translation	13,366
RWN base concepts manual translation	135
Hypernyms manual translation	3,661

The development of TatThes was carried out by the following ways: 1) Supplementing an existing RuThes concept with Tatar lexical entries. Due to the language-independent nature of the RuThes conceptual network, it is mostly reused in Tatar thesaurus, even though the RuThes concepts can be expressed in Russian and Tatar texts in very different ways. For example, the *Age of majority* concept is expressed in Russian by the one noun "совершеннолетие", but in Tatar it is expressed by three verb phrases "буйга җитү", "яше җитү" and "балигъ булу".

A reused RuThes concept is supplemented by the Tatar translation of the concept name and by Tatar lexical entries.

The Russian and the Tatar lexical entries of the same concept can be described as cross-lingual ontological synonyms.

2) Adding a new hyponym concept and its lexical entries. The RuThes conceptual network can lack the concepts, specific to socio-cultural life of the Tatar society, such as Islam-related notions, social hierarchy of Oriental societies, Tatar ethno-cultural phenomena, etc. Such the concepts were added to TatThes as hyponyms of the existing RuThes concepts. For example, the *Muslim holiday* concept was added as a narrower concept of the *Holiday* concept.

3) Adding a new intermediate concept and its lexical entries. Even though the RuThes conceptual network is language-independent, it is nevertheless linguistically-motivated and thus can lack the concepts, lexicalized in Tatar, but not lexicalized in Russian. Many of such the concepts were added to TatThes on the intermediate level of the conceptual network, i.e. as a hyponym of one concepts and as a hypernym of the another. For example, RuThes contains *Stepson* and *Stepdaughter* concepts, but doesn't contain the concept of *Stepchild*. This concept was added to TatThes as a hyponym concept of the Relative and a hypernym of the *Stepson* and *Stepdaughter* concepts.

TatThes has been published on the Linguistic Linked Open Data cloud as part of RuThes Cloud project [12].

TatWordNet

With the described resources in hands, we constructed TatWordNet by the following steps: 1) Semi-automatic conversion of TatThes concepts to TatWordNet synsets. This conversion was performed in the same way as conversion of RuThes to RuWordNet. 2) Semi-automatic translation of the RuWordNet concepts. At first, we automatically translated the lexical entries by the Ganiev bilingual Russian-Tatar dictionary. Then we manually filtered out the incorrect translations, adding correct variants where necessary. 3) Manual translation of the Base RuWordNet concepts. 4) Manual translation of the hypernyms and holonyms of the previously translated RuWordNet concepts.

The number of TatWordNet synsets, obtained on each step is represented at Table 1.

 4   TatWordNet description

TatWordNet is organized as networks of synsets, where each synset is linked to its lexical entries via lexical senses. The resource is distributed under the Creative Commons Attribution-ShareAlike License.

Linked Open Data representation

TatWordNet has been published to the Linguistic Linked Open Data cloud [6] and interlinked with the Global WordNet Grid [21] via the Collaborative Interlingual Index [4].

The resource is represented in terms of Global WordNet ontology as well as Onto-Lex/Lemon, SKOS, LexInfo and PROV ontologies.

Listing 1 represents the *City* synset and one of its lexical entries, senses and synset relations.

 Listing 1 The *City* synset, its lexical entries, senses and relations.

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix wn: <https://globalwordnet.github.io/schemas/wn#>.
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#>.
@prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>.
@prefix prov: <http://www.w3.org/ns/prov#>.

#Synset City
<http://lod.wordnet.tatar/synset/242-N>
  a ontolex:LexicalConcept , skos:Concept;
  wn:partOfSpeech wn:noun;
  skos:altLabel "шәһәр"@tt , "кала"@tt;
  ontolex:isEvokedBy
    <http://lod.wordnet.tatar/entry/шәһәр>,
    <http://lod.wordnet.tatar/entry/кала>;
  ontolex:lexicalizedSense
    <http://lod.wordnet.tatar/sense/242-N-шәһәр>,
    <http://lod.wordnet.tatar/sense/242-N-кала>;
  wn:domain_topic
    <http://lod.wordnet.tatar/synset/1702-N>;
  wn:hypernym
    <http://lod.wordnet.tatar/synset/123680-N>,
    <http://lod.wordnet.tatar/synset/145516-N>;
  wn:hyponym
    <http://lod.wordnet.tatar/synset/207-N>,
    <http://lod.wordnet.tatar/synset/3208-N>,
    <...>;
  wn:mero_part
    <http://lod.wordnet.tatar/synset/9171-N>,
    <http://lod.wordnet.tatar/synset/4250-N>,
    <http://lod.wordnet.tatar/synset/4773-N>;
  skos:inScheme
    <http://lod.wordnet.tatar/tatwordnet>;
  prov:wasGeneratedBy
    <http://lod.wordnet.tatar/prov/ganiev-translation>.

#Synset relation
<http://lod.wordnet.tatar/relation/hypernym-from-242-N-to-123680-N>
```

```
  a vartrans:LexicoSemanticRelation ;
  vartrans:category wn:hypernym;
  vartrans:source
    <http://lod.wordnet.tatar/synset/242-N>;
  vartrans:target
    <http://lod.wordnet.tatar/synset/123680-N>.

#Lexical sense, linking the concept to its lexical entry
<http://lod.wordnet.tatar/sense/242-N-шәһәр>
  a ontolex:LexicalSense ;
  ontolex:reference
    <http://lod.wordnet.tatar/synset/242-N>;
  ontolex:isLexicalizedSenseOf
    <http://lod.wordnet.tatar/synset/242-N>;
  ontolex:isSenseOf
    <http://lod.wordnet.tatar/entry/шәһәр>.

#Lexical entry
<http://lod.wordnet.tatar/entry/шәһәр>
  a ontolex:LexicalEntry, ontolex:Word;
  rdfs:label "шәһәр"@tt;
  wn:partOfSpeech wn:noun;
  lexinfo:partOfSpeech lexinfo:noun;
  ontolex:canonicalForm
    [
    rdf:type ontolex:Form;
    ontolex:writtenRep "шәһәр"@tt;
    ];
  ontolex:evokes
    <http://lod.wordnet.tatar/synset/242-N>;
  ontolex:sense
    <http://lod.wordnet.tatar/sense/242-N-шәһәр>;
  prov:wasGeneratedBy
    <http://lod.wordnet.tatar/prov/ganiev-translation>.
```

This representation is quite straightforward and mainly reflects GWA recommendations. The only three comments should be made: (1) In order to achieve computability with standard SKOS applications, the lexical entries were additionally represented as SKOS labels. (2) In accordance with OWL punning, the synset relationships were represented as RDF object properties and at the same time as RDF individuals. (3) Each synset was provided by the `prov:wasGeneratedBy` link to the method used to produce this synset (i.e. automatic generation from a TatThes concept, manual translation of a RuWordNet synset, etc).

Publishing on the Web

TatWordNet has been published on the Web (`http://wordnet.tatar/`) and is available via:
- dereferenceable URIs: `http://lod.wordnet.tatar`;
- SPARQL endpoint: `http://lod.wordnet.tatar/sparql`;
- RDF dump: `http://wordnet.tatar/download/twn.ttl.zip`.

Access to the resource via dereferenceable URIs is supported by mechanisms of content negotiation. When a web browser requests a URI, it is redirected to a web page with an HTML view of the entity, but the Semantic Web agent request is redirected to the page with the RDF representation.

■ **Table 2** TatWordNet statistics.

Entity type	Count
Synset	18,538
Lexical entry	36,540
Word	13,469
Multi-word expression	23,071
Lexical sense	49,525
Synset relation	68,558
hypernym / hyponym	24,740
instance hypernym / hyponym	221
part holonym / meronym	1,336
domain topic	15,964
Link to inter-lingual index	3,661

Statistics

Statistics of RuThes Cloud is represented at Table 2.

SPARQL query example

Integration of TatWordNet to the LLOD cloud makes it possible to construct very complex federated SPARQL queries. Example of such queries is the following: find the Russian sentences, containing the words whose Tatar translations are hyponyms of the given Tatar word "шәһәр" ("city") (Listing 2). This query utilizes several types of links: (1) between Russian corpora OpenCorpora and RuThes thesaurus, (2) cross-lingual links between RuThes and TatWordNet, and (3) finally internal links between TatWordNet synsets.

■ **Listing 2** SPARQL query example.

```
PREFIX wn: <http://globalwordnet.github.io/schemas/wn#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex#>
PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
PREFIX conll: <http://ufal.mff.cuni.cz/conll2009-st/task-description.html#>

SELECT ?twn_hyponym ?opencorpora_word_text ?opencorpora_sentence_text {
    "шәһәр"@tt
        ^ontolex:writtenRep /
        ^ontolex:canonicalForm /
        ^ontolex:isEvokedBy ?twn_synset.
    ?twn_synset (wn:hyponym | wn:instance_hyponym)* ?twn_hyponym.
    ?twn_hyponym skos:closeMatch ?ruthes_concept.

    SERVICE <http://data.llod.ru/repositories/ruthes-cloud> {
        ?ruthes_concept
            ontolex:lexicalizedSense /
            ontolex:isSenseOf ?ruthes_entry.

        SERVICE <http://data.llod.ru/repositories/opencorpora> {
            ?ruthes_entry ^conll:LEXICAL_ENTRY_URI ?opencorpora_word.
            ?opencorpora_word nif:sentence ?opencorpora_sentence.
            ?opencorpora_word nif:anchorOf ?opencorpora_word_text.
            ?opencorpora_sentence nif:anchorOf ?opencorpora_sentence_text.
        }
    }
}
```

The query results contain 113 sentences, for example "Российскую <u>столицу</u> впервые посетил известнейший художник-визионёр Алекс Грей" ("The famous painter Alex Grey made his first visit to the Russian <u>capital</u>").

5 Conclusion

In this paper, we present the first release of TatWordNet (`http://wordnet.tatar`), a Linguistic Linked Open Data-integrated wordnet resource for Tatar. TatWordNet was constructed on the base of three source resources, developed by us: RuThes, RuWordNet and TatThes. The currents version of TatWordNet contains 18,583 synsets, 36,540 lexical entries and 49,525 senses. The resource has been published to the Linguistic Linked Open Data cloud and interlinked with the Global WordNet Grid.

References

1　Özge Bakay, Özlem Ergelen, Elif Sarmış, Selin Yıldırım, Atilla Kocabalcıoglu, Bilge Nas Arıcan, Merve Özçelik, Ezgi Sanıyar, Oguzhan Kuyrukcu, Begüm Avar, and Olcay Yildiz. Turkish WordNet KeNet. In Piek Vossen and Christiane Fellbaum, editors, *Proceedings of the 11th Global Wordnet Conference (GWC 2021), Potchefstroom, South Africa, 18–21 Jan, 2021*, pages 166–174. GWA, 2021. URL: `https://www.aclweb.org/anthology/2021.gwc-1.19/`.

2　Orhan Bilgin, Özlem Çetinoğlu, and Kemal Oflazer. Building a Wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1–2):163–172, 2004. URL: `http://research.sabanciuniv.edu/379/`.

3　Francis Bond and Ryan Foster. Linking and Extending an Open Multilingual Wordnet. In Hinrich Schuetze et al., editors, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, 4–9 August, 2013. Volume 1: Long Papers*, pages 1352–1362. ACL, 2013. URL: `https://www.aclweb.org/anthology/P13-1133`.

4　Francis Bond, Piek Vossen, John P. McCrae, and Christiane Fellbaum. CILI: The collaborative interlingual index. In Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, and Corina Forascu, editors, *Proceedings of the 8th Global WordNet Conference (GWC 2016), Bucharest, Romania, 27–30 January, 2016*, pages 50–57. GWA, 2016. URL: `https://www.aclweb.org/anthology/2016.gwc-1.9/`.

5　Özlem Çetinoğlu, Orhan Bilgin, and Kemal Oflazer. Turkish Wordnet. In Kemal Oflazer and Murat Saraçlar, editors, *Turkish Natural Language Processing*, pages 317–336. Springer, 2018. `doi:10.1007/978-3-319-90165-7_15`.

6　Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia. Linguistic Linked Open Data Cloud. In Philipp Cimiano et al., editors, *Linguistic Linked Data: Representation, Generation and Applications*, pages 29–41. Springer, 2020. `doi:10.1007/978-3-030-30225-2_3`.

7　Elin Ehsani. *KeNet: A Comprehensive Turkish Wordnet and Using it in Text Clustering*. PhD thesis, Işık University, 2018. `doi:10.13140/RG.2.2.20932.27524`.

8　Razieh Ehsani, Ercan Solak, and Olcay Taner Yildiz. Constructing a WordNet for Turkish Using Manual and Automatic Annotation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 17(3), April 2018. `doi:10.1145/3185664`.

9　Christiane Fellbaum. Wordnet. In Roberto Poli et al., editors, *Theory and Applications of Ontology: Computer Applications*, pages 231–243. Springer, 2010. `doi:10.1007/978-90-481-8847-5_10`.

10　Alfiya Galieva, Alexander Kirillovich, Bulat Khakimov, Natalia Loukachevitch, Olga Nevzorova, and Dzhavdet Suleymanov. Toward Domain-Specific Russian-Tatar Thesaurus Construction. In Radomir Bolgov et al., editors, *Proceedings of the International Conference IMS-2017, St. Petersburg, Russia, 21–24 June 2017*, ACM International Conference Proceeding Series, pages 120–124. ACM Press, New York, 2017. `doi:10.1145/3143699.3143716`.

11　Adam Kilgarriff and Christiane Fellbaum. *WordNet: an Electronic Lexical Database*. MIT Press, 2000.

12　Alexander Kirillovich, Olga Nevzorova, Emil Gimadiev, and Natalia Loukachevitch. RuThes Cloud: Towards a Multilevel Linguistic Linked Open Data Resource for Russian. In Przemysław

Różewski and Christoph Lange, editors, *Proceedings of the 8th International Conference on Knowledge Engineering and Semantic Web (KESW 2017), Szczecin, Poland, November 8–10, 2017*, Communications in Computer and Information Science, vol. 786, pages 38–52. Springer, 2017. `doi:10.1007/978-3-319-69548-8_4`.

13 N. Loukachevitch, B. Dobrov, and I. Chetviorkin. RuThes-lite, a Publicly Available Version of Thesaurus of Russian Language RuThes. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue»*, pages 340–349. RGGU, 2014. URL: `http://www.dialog-21.ru/digests/dialog2014/materials/pdf/LoukachevitchNV.pdf`.

14 N. V. Loukachevitch, G. Lashevich, A. A. Gerasimova, V. V. Ivanov, and B. V. Dobrov. Creating Russian WordNet by Conversion. In *Computational Linguistics and Intellectual Technologies: papers from the Annual Conference "Dialogue"*, pages 405–415. RGGU, 2016. URL: `http://www.dialog-21.ru/media/3409/loukachevitchnvetal.pdf`.

15 Natalia Loukachevitch and Boris Dobrov. RuThes Linguistic Ontology vs. Russian Wordnets. In Heili Orav, Christiane Fellbaum, and Piek Vossen, editors, *Proceedings of the 7th Global Wordnet Conference (GWC 2014), Tartu, Estonia, 25–29 January, 2014*, pages 154–162. University of Tartu Press, 2014. URL: `https://www.aclweb.org/anthology/W14-0121/`.

16 Natalia Loukachevitch and Boris Dobrov. RuThes Thesaurus for Natural Language Processing. In Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev, editors, *The Palgrave Handbook of Digital Russia Studies*, pages 319–334. Palgrave Macmillan, 2021. `doi:10.1007/978-3-030-42855-6_18`.

17 Natalia Loukachevitch, German Lashevich, and Boris Dobrov. Comparing two thesaurus representations for Russian. In Francis Bond, Takayuki Kuribayashi, Christiane Fellbaum, and Piek Vossen, editors, *Proceedings of the 9th Global Wordnet Conference (GWC 2018), Singapore, 8–12 January, 2018*, pages 34–43. GWA, 2018. URL: `https://www.aclweb.org/anthology/2018.gwc-1.5/`.

18 Roberto Navigli and Simone Paolo Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, December 2012. `doi:10.1016/j.artint.2012.07.001`.

19 D. Tufis, D. Cristeau, and S. Stamou. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal of Information Science and Technology*, 7(1–2):9–43, 2004. URL: `http://www.dblab.upatras.gr/balkanet/journal/7_Overview.pdf`.

20 Piek Vossen. EuroWordNet General Document. Technical Report LE2-4003, LE4-8328, University of Amsterdam, July 1999. URL: `http://www.illc.uva.nl/EuroWordNet/docs.html`.

21 Piek Vossen, Francis Bond, and John P. McCrae. Toward a truly multilingual global wordnet grid. In Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, and Corina Forascu, editors, *Proceedings of the 8th Global WordNet Conference (GWC 2016), Bucharest, Romania, 27–30 January, 2016*, pages 419–426. GWA, 2016. URL: `https://www.aclweb.org/anthology/2016.gwc-1.59/`.