მონაცემის გაცვლის ენები

დავით ურიდია

Email: uridiadato@gmail.com

**რეზიუმე** — მსოფლიო ქსელში (World Wide Web) XML პოპულარობა რამოდენიმე მიზეზის გამო მოიპოვა. პირველ რიგში, მჭიდრო კავშირშია წინამორბედ სტანდარტთან Standard Generalized Markup Language (SGML) და Hypertext Markup Language (HTML), ამ ორ ენაზე დაწერილი ინფორმაცია მარტივად ფართოვდება XML-ზე. მაგალითისთვის, XHTML განისაზღვრა ფორმატში, რომელშიც არის ფორმალური XML. ამან გამოიწვია ვებ ბრაუზერებში XML მხარდაჭერა და toochains გამოიყენება ვებ ფეიჯების გენერირებისთვის. XML-ის დიზაინის მიზნები ხაზს უსვამს ინფორმაციის გადაცემის სიმარტივეს ინტერნეტსა და მისი გამოყენების შესაძლებლობას. ეს არის ტექსტური მონაცემის ფორმატი Unicode-ს მხარდაჭერით. მიუხედავათ ამისა, XML-ის დიზაინი ფოკუსირებულია დოკუმენტებზე, ენა კი ფართოდ გამოიყენება ვებ სერვისებში.

**გასაღები სიტვვები -** Data Exchange Languages, XML, www(World Wide Web)

2017-2018 სასწ.წელი თსუ კომპიუტერულ მეცნიერებათა დეპარტამენტი

—————————— ◆ ——————————

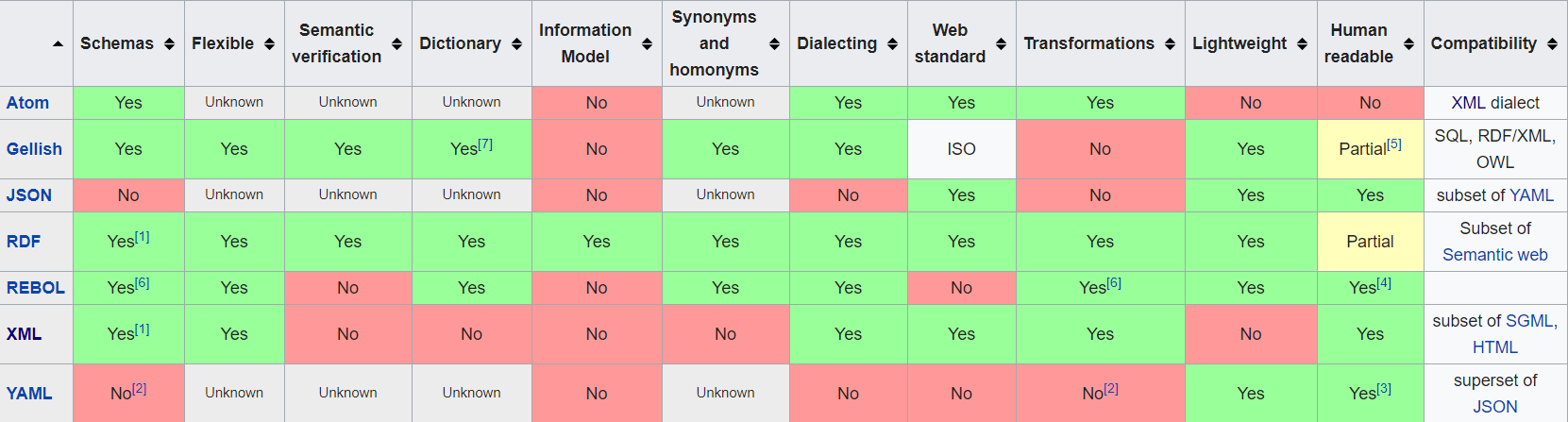
# 1 შესავალი

ქსელში ინფორმაცია მიმოცვლა საშუალებას აძლევს EPA და tribes გაცვალონ მონაცემი წინასწარ განსაზღვრული პროტოკოლებისა და სტანდარტების ნაკრების გამოყენებით. ქსელში გაცვლილ ნებისმიერი სახის ინფორმაცია არის მონაცემთა გაცვლა(data exchange). მონაცემის გაცვლას ქსელში გაცვლის(Exchange Network) დისკუსიებსში ხშირად მოიხსენიებენ როგორც ნაკადებს (flows) ან ინფორმაციის ნაკადებს (data flows). მონაცემის გაცვლა არის სპეციფიური ტიპის ინფორმაციის გაცვლა ორ ან მეტ პარტნიორს შორის. გაცვლებს ხშირად ერთი ინტერესის ქვეშ აერთიანებენ, EPA პროგრამა (როგირცაა the Toxic Release Inventory) ან EPA მონაცემთა ბაზა (როგორიცაა RCRAInfo).

მონაცემთა გაცვლის ენა არის ენა, რომელიც არის დომეინურ-დამოუკიდებელი და შეიძლება გამოყენებულ იქნას ნებისმიერი სახის მონაცემებთან. მისი შესაძლებლობებისა და ხარისხის სემანტიკური გამოხატულება არის ფარდოთ განსაზღვრული ბუნებრივი ენის შესაძლებლობასთან შედარებასთან. ტერმინი ასევე ვრცელდება ნებისმიერი ფაილის ფორმატზე, რომელსაც კითხულობს ერთზე მეტი პროგრამა, მათ შორის Microsoft Office დოკუმენტიც. თუმცა, ფაილის ფორმატი არ არის რეალური ენა რადგან მას აკლია გრამატიკა და ლექსიკა.

# 2 განხორციელებული სამუშაო (არსებული მდგომარეობა)

პოპულარული ენებია: RDF, XML, Atom, JSON, YAML, REBOL, Gelilish..



**ნომეკლატურა**

* Schema (სქემა) - არის თუ არა ენის განმარტება კომპიუტერულ-ინტერპრეტაციურ ფორმაში.
* Flexible (მოქნილი) - ენა საშუალებას აძლევს სქემას გაფართოვდეს მისი მოდიფიკაციის გარეშე.
* Semantic verification (სემანტიკური ვერიფიკაცია) - ენის სისწორე.
* Dictionaty - მემკვიდრეობით ენა შეიცავს ლექსიკონსა და ტოპოლოგიებს.
* Synonyms&homonyms (სინონიმები&ომონიმები) - ენა შეიცავს სინონიმებსა და ომონიმებს.
* Dialecting (დიალექტი) - იქნება თუ არა ენის განმარტება მრავალ ენაზე ან დიალექტზე.
* Web or ISO სტანდარტი - ენები რეკომენდირებული როგორც სტანდარტი.
* Transformations (ტრანსფორმაცია) - შეძლებს თუ არა ენა სხვა სტანდარტზე გარდაქმნას.
* Lightweigh (მსუბუქი) - რამდენად მსუბუქი ვერისაა ხელმისაწვდომი, სრული ვერსიის გარდა.
* Human readable (წაკითხვადი ადამიანისთვის) - წაკითხვადი ნებისმიერი ადამიანისთვის.
* Compatibility (თავსებადობა) - სხვა რომელი ინსტრუმენტების გამოყენებაა შესაძლებელი ან საჭირო ამ ენის გამოყენების დროს.

**შენიშვნები**

1. RDF-ის სქემა არის მოქნილი ენა.
2. XML სქემა შეიცავს კიმიტირებულ გრამატიკასა და ლექსიკას.
3. YAML ხელმისაწვდომია როგორც გაფართოება.
4. REBOL დეფაულტ ფორმატია და არა კომპაქტური სინტაქსი.
5. Gellish სინტაქსი ძალიან მარტივია, მაგრამ დიალექტებს სჭირდება დომეინური ცოდნა.

# 3 სისტემური მოდელი

Extensible Markup Language (XML) არის მარკირებული ენა, რომელიც განსაზღვრავს წესების ნაკრებს დოკუმენტის კოდირებისთვის და ეს დოკუმენტი წაკითხვადია როგორც ადამიანისთვის, ისე მანქანისთვის. შემუშავდა ასობით დოკუმენტის ფორმატი XML სინტაქსის გამოყენებით, მათ შრის RSS, Atom, SOAP, SVG და XHTML. XML-ზე დაფუძნებული ფორმატები გახდნენ ბევრი საოფისე პროდუქტის ინსტრუმენტები, მათ შორის Microsoft Office, OpenOffice.org, LibreOffice და ეფლის iWork. XML უზრუნველყო ბაზის ურთიერთობისთვის პროტოკოლი XMPP. Microsoft-ის .NET Framework კონფიგურაციისთვის იყენებს XML ფაილს. Apple გააჩნია რეგისტრზე დაფუძნებული XML . XML სპეციფიკაცია შემდეგნაირია.

**სიმბოლო**

XML დოკუმნეტი არის სიმბოლოების სტრინგი. თითქმის ყველა Unicode ტიპის სიმბოლო გამოჩნდება XML დოკუმენტში.

**პროცესორი და აპლიკაცია**

პროცესორი აანალიზებს მარკირებას და აგზავნის სტრუქტურირებულ ინფორმაციას განაცხადზე.

**მარკირება და შინაარსი**

XML დოკუმენტის შემქმნელი სიმბოლოები იყოფა მარკირებისა და შინაარის სიმბოლოებად, რომლებიც გამოირჩევიან მარტივი სინტაქსური წესების გამოყენებით. ზოგადად, სტრინგი იწყება < სიმბოლოთ და მთავრდება >, ან იწყება & და მთავრდება ;. სიმბოლოების სტრინგი რომელიც მარკირებული არაა შინაარსია. თუმცა CDATA სექციაში <![CDATA[ and ]]> კლასიფიცირებულია როგორც მარკირება, მაშინაც კი როდესაც მის შიგნით არსებული ტექსტი შინაარსია.

**Tag**

თაგი არის მარკირების კონსტრუქცია რომელიც იწყება < და მთავრდება >. სამი ძირითადი თაგია:  
start-tag, როგორიცაა <section>;

end-tag, როგორიცაა </section>;

empty-element tag, როგორიცაა <line-break />;

**ელემენტი**

ელემენტი არის ლოგიკური დოკუმენტის კომპონენტი, რომელიც იწყება start-tag და მთავრდება end-tag-ით ან შეიცავს მხოლოდ empy-element tag. სიმბოლოები start-tag-სად და end-tag-ს შორის არის ელემენტი და შეიძლება იყოს მარკირებული სხვა ელემენტებთან ერთად, რომელსაც შვილობილ ელემენტებს უწოდებენ. მაგ; <greeting>Hello, world!</greeting>.  <line-break />.

**ატრიბუტები**

ატრიბუტი წარმოადგენს მარკირებას შედგარი name-value წყვილისგან, რომელიც არსებობს start-tag-სა და empty-tag-თან ერთად. მაგალითისთვის <img src="madonna.jpg" alt="Madonna" />, სადაც სახელის ატრიბუტებია src და alt და მათი მნიშვნელობებია “madonna.jpg” და “Madona”. შემდეგი მაგალითია <step number="3">Connect A to B.</step>, ატრიბუტია “number” და მისი მიშვნელობაა “3”. XML ატრიბუტს აქვს მხოლოდ ერთი მნიშვნელობა და თითოეულს ატრიბუტს შეუძლია გამოჩნდეს მინიმუმ ერთხელ ერთ ელემენტზე. ზოგ სიტუაციაში შესაძლებელია მნიშვნელობათა სიის შექმნა, კარგად ფორმირებული XML ატრიბუტების კოდირებით. როგორც წესი ელემენტებს გამოყოფენ , ; ან ჰარით. <div class="inner greeting-box">Welcome!</div> ატრიბუტი კლასი შეიცავს ორივე მნიშვნელობას, როგორც “inner” ასევე „greeting-box“.

**XML დეკლარაცია**

XML დოკუმენტები შეიძლება დაიწყოს XML-ის დეკლარაციით, რომელიც აღწერს მათ შესახებ გარკვეულ ინფორმაციას. <?xml version="1.0" encoding="UTF-8"?>

# 4 სქემა და ვალიდაცია

იმისათვის რომ კარგად იყოს ჩამოყალიბებული XML დოკუმენტი შეიძლება იყოს ძალაში. ეს ნიშნავს, იმას რომ იგი შეიცავს რეფერენს Document Type Definition (DTD) და მასში არსებული ელემენტები და ატრიბუტები იცავს გრამატიკულ წესებს, რომელიც განსაზღვრავს DTD-ს.

DTD არის მაგალითი სქმისა ან გრამატიკის. XML 1.0 გამოქვეყნების შემდეგ სქემის ენაზე მნიშვნოლავანი სამუშაო ჩატარდა. ასეთი სქემების ენები, როგორც წესი, ზღუდავენ იმ ელემენტების კომპლექტს, რომლებიც გამოყენებულია დოკუმენტში, რომელთა ატრიბუტები შეიძლება იყოს მათთვის ბრძანება, რომელიც შეიძლება გამოჩნდეს და დასაშვებია მშობელი / ბავშვის ურთიერთობებში.

ახალი სქემა აღწერილი W3C მიერ, არის XML სქემა რომელსაც ხშირად მოიხსენიებენ XSD(XML Schema Definition). XSD არის გაცილებით ძლიერი ვიდრე DTD. იყენებს მდიდარ datatyping-ს და იძლევა უფრო დეტალურ შესღუდვებს XML კოდუმენტის ლოგიკურ სტრუქტურაში. XSD ასევე იყენებს XML- ზე დაფუძნებულ ფორმატს, რაც საშუალებას იძლევა გამოიყენონ ჩვეულებრივი XML ინსტრუმენტები, მათ დასახმარებლად.

# დასკვნა

პრაქტიკა გვიჩვენებს, რომ ფორმალური ენების გარკვეული ტიპები უფრო შესაფერისია ასეთი ამოცანებისათვის ვიდრე სხვები, რადგან სპეციფიკა მუშაობს ფორმალური პროცესებით და არა კონკრეტული პროგრამული უზრუნველყოფით. მაგალითად, XML არის მარკირებული ენა, რომელიც მიზნად ისახავს დიალექტების შექმნას(სპეციფიკური დომეინის ქვეენა) და არის პოპულარული არჩევანი ინტერნეტში. თუმცა, ის არ შეიცავს დომეინის სპეციფიურ ლექსიკონებსა და ფაქტორებს.

**ლიტერატურა**

1. R. Fagin, P. Kolaitis, R. Miller, and L. Popa. "[Data exchange: semantics and query answering](http://www.sciencedirect.com/science/article/pii/S030439750400725X)." Theoretical Computer Science, 336(1):89–124, 2005W.-K. Chen, *Linear Networks and Systems.* Belmont, Calif.: Wadsworth, pp. 123-135, 1993. (Book style)
2. P. Kolaitis. "[Schema mappings, data exchange, and metadata management](http://se-pubs.dbs.uni-leipzig.de/files/Kolaitis2005Schemamappingsdataexchange.pdf)." Proceedings of the twenty- fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 61–75, 2005
3. ["XML Media Types, RFC 7303"](http://www.rfc-editor.org/rfc/rfc7303.txt). Internet Engineering Task Force. July 2014.
4. [XML 1.0 Specification"](http://www.w3.org/TR/REC-xml). World Wide Web Consortium. Retrieved 22 August 2010.
5. https://www.google.ge/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjq6O-mhJ7YAhWKKJoKHU1MDSoQFggtMAE&url=http%3A%2F%2Fwww.exchangenetwork.net%2Fdev\_schema%2Fexchange\_design\_guidance\_v1.2.doc&usg=