

# Семантик Вэб Технологийг Цахим Номын Сангийн Системд Нэвтрүүлэх

Н.Саранчимэг

Мэдээллийн системийн менежментийн тэнхим  
Санхүү эдийн засгийн дээд сургууль  
Улаанбаатар, Монгол  
[saranchimeg@ife.edu.mn](mailto:saranchimeg@ife.edu.mn)

## Товч хураангуй

Интернэт вэб технологийн хөгжлийн ачаар бид мэдээллийн эрин зуун буюу мэдээллийн хөлгүй их далайн гүнд амьдрах болжээ. Хэдий бидэнд олон зун вэб нөөцүүд, хайлтын системүүд байгаа ч тэдгээр нь бидэнд хэрэгцээтэй, шаардлагатай мэдээллийг олоход төдийлөн оновчтой туслалцаа үзүүлж чадахгүй байна. Google гэх мэт хайлтын томоохон системүүд бидний мэдээллийн эрэл хайгуулд тодорхой дэмжлэг өгч байгаа ч түлхүүр үгэнд суурилсан хайлтын хэдэн мянган магадгүй хэдэн сая илэрцээс өөрсдийн хэрэглэх мэдээллийг олоход хамгийн төвөгтэй хэцүү ажил хэвээр л байна.

Манайх шиг хөгжиж байгаа орны хувьд мэдлэгийг бий болгох бүтцийг тодорхойлж эрдэм шинжилгээ судалгааны салбартаа оновчтой шийдлийг хөгжүүлэх, түүгээрээ дамжуулан оюутан, багш эрдэмтэн судлаачдад суралцах, судлах нөөц бололцоог нь бүрдүүлэхийн тулд номын сангийн системд шинэлэг, дэвшилтэт шийдэл бий болгон хөгжүүлэх хэрэгтэй юм.

Сүүлийн жилүүдэд вэб технологи дараагийн шатны хөгжүүлэлт буюу семантик вэб технологийг боловсруулан бидэнд илүү ухаалаг харну өгч чадах вэбүүдийг бүтээх ажил өрнөсөөр байна. Нэн ялангуяа хамгийн үнэт мэдээллийн сан буюу номын сангийн системд семантик вэбийн технологийг нэвтрүүлэх талаар олон төсөл хөтөлбөрүүд хэрэгжиж байна.

Түлхүүр үг: *Онтологи; Метадата; RDF схем; Цахим номын сан;*

## I. Үндсэн Асуудал

Манай улсад одоогоор 1000 гаруй номын сан үйл ажиллагаа явуулж байгаагаас судалгаа шинжилгээний материалууд хадгалагддаг 200 орчим их дээд сургуулийн, 50 орчим эрдэм шинжилгээний болон албан байгууллагын тусгай номын сангууд байдаг [6]. Гол асуудал нь эдгээр номын санд хадгалагдаж байгаа судалгааны ажлуудтай танилцах судлахын тулд заавал биеэр очиж үйлчлүүлэх шаардлагатай байдаг. Хэдийгээр

их дээд сургуулиуд болон нийтийн номын сангууд цахим номын сангийн програм хангамж хэрэглэдэг хэдий ч зарим сайжруулалтуудыг хийх шаардлагатай гэж үзэж байна.

- Тухайн эрдэм шинжилгээний судалгааны ажлуудын хураангуй /abstract/ хувилбарыг нээлттэй байлгах. Мөн бүтэн бичвэрийг тодорхой төлбөртэйгөөр цахим хэлбэрээр ашиглах боломжийг бүрдүүлэх
- Номын сангийн хэрэглэгчдийн хамтын ажиллагааг сайжруулан мэдээллээ хуваалцах туршлагаа солилцох орчинг бүрдүүлэх
- Тухайн нэгдсэн системд семантик вэб технологийг нэвтрүүлэн мэдээллийн илэрцийн чанарыг сайжруулан мэдлэг бүтээх процессийг боловсронгуй болгох

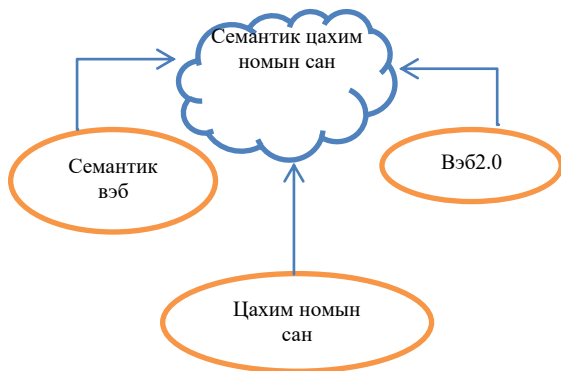
Маш олон тооны төсөл хөтөлбөрүүд судалгаа шинжилгээний ажлууд боловсруулагдаж байгаа хэдий ч нийтэд төдийлөн нээлттэй биш, мэдлэгээ хуваалцах, хамтран ажиллах нэгнээсээ суралцах боломж бүрдэхгүй байна. Мөн цахим орчинд тавигдаж байгаа мэдээлүүд төдийлөн баталгаатай, ул үндэстэй мэдээлэл байдаггүй бөгөөд хамгийн чанартай, үнэ цэнэтэй мэдээллийг бий болгохын тулд номын сангийн нэгдсэн сүлжээг боловсруулан оновчтой мэдээллийг олох боломж бүрдүүлэхийн тулд семантик вэб технологийг ашиглах хэрэгтэй юм.

W3C (*World Wide Web Consortium*) сүүлийн жилүүдэд семантик технологийг вэб орчинд хөгжүүлэх тал дээр маш идэвхтэй судалгаа шинжилгээ төсөл хөтөлбөрүүд хэрэгжүүлж байна. Эдгээрийн нэг гол төсөл нь номын сангийн системүүдэд семантик технологийг нэвтрүүлэх төсөл хэрэгжүүлж эхлээд байгаа юм. Жишээ нь: Talia бол хүмүүнлэгийн салбарын судалгааны ажлуудын цахим сан бөгөөд семантик технологид суурилан хөгжүүлэлт хийгдэж байна. Энэхүү систем нь мэдлэгийн зохион байгуулалт хийхдээ онтологийг ашиглаж байгаа бөгөөд судалгаа шинжилгээний ухаалаг орчинг бий болгох

зорилготой ажиллаж байна.[1] Мөн JeromeDL<sup>1</sup> семантик цахим номын сан нь нээлттэй эх бүхий амжилттай хэрэгжиж байгаа төсөл бөгөөд олон нийтийн сүлжээний технологийг ашиглан мэдээллийн нөөцийн ашиглалт болон чанарыг нэмэгдүүлэх боломжуудыг тусгаж өгсөн байна.

## II. ТЕХНОЛОГИ

Семантик цахим номын сангийн шийдлийг хэрэгжүүлэхэд **семантик вэб** болон **Вэб2.0** технологиудыг ашиглан номын сангийн нөөцүүдээс хайлт хийж хүссэн мэдээллээ олж авах болон нийтийн сүлжээний хамтын ажиллагааны технологийн шийдлийг тусгаж өгснөөр илүү ач холбогдолтой, үр өгөөжтэй болох юм. Зур.1-т бүтцийн бүдүүвч зургийг харууллаа.

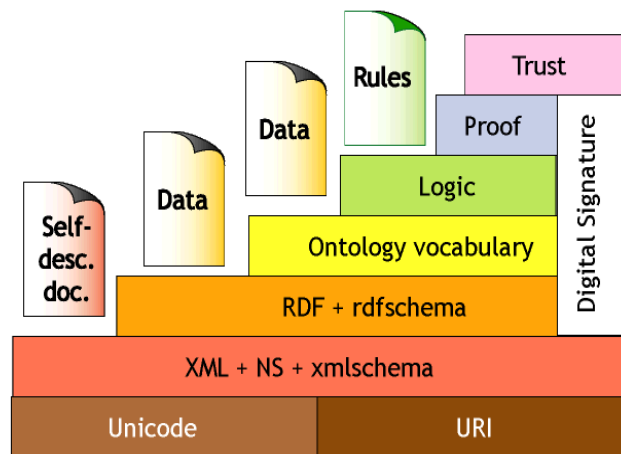


Зураг 1. Семантик цахим номын сангийн бүтэц  
Эх сурвалж: Sebastian Ryszard Kruk- Semantic Digital Libraries

### A. Семантик вэб

Вэб дээрх мэдээллийг илүү мэдлэгийн шинж чанартай болгон хөгжүүлэхийн тулд мэдээллийг тодорхой хэлбэрт оруулах бөгөөд хүн болон машин /компьютер/ аль нь ч ойлгох боломжтой байдлаар бүтэцлэх технологи. Семантик вэб дээр яригддаг гол ойлголтуудын нэг бол онтологи буюу аливаа юмсын учир шалтгаан болон хоорондын уялдаа холбоог тодорхойлж өгөх явдал юм. Философийн шинжлэх ухаанд яригддаг энэхүү онтологийн ойлголтыг вэб технологид шингээж өгснөөр вэб өөрийн агуулгыг ойлгож боловсруулах чадвартай болох юм. Зарчмын хувьд маш энгийн бөгөөд аливаа зүйл хоорондоо уялдаа холбоотойгоор илэрхийлэгддэг гэсэн зарчмыг RDF (*Resource Description Framework*) хэлээр вэб дээрх мэдээллийг бүтэцлэж дүрслэх юм. Ерөнхийдөө RDF нь онтологийн зарчимд тулгуурласан вэбийн өгөгдөл солилцооны

загварчлал бөгөөд Зур.2-т дүрсэлсэн Объект-Шинж чанар-Утга гэсэн мэдэгдэл (statement) гэж нэрлэгдэх тулгуур элемент дээр суурилсан байдаг. RDF нь бие даан ямар нэгэн үйл ажиллагааг дүрслэж чадахгүй бөгөөд RDFs схемийн тусламжтайгаар утга санаагаа хэрэгжүүлдэг [2]. RDF-ээр хэлбэржүүлсэн бүтцээр хадгалагдсан өгөгдөлтэй ажиллах, мэдээллийг боловсруулахын тулд SPARQL (*Sparkle Protocol and RDF Query Language*) эрэлт хүсэлтийн хэлийг ашигладаг. Харин OWL (*Web ontology language*) нь объектууд хоорондын хамаарлыг тодорхойлж өгсөн дүрэм журмын цогцыг загварчилсан хэл юм. Өөрөөр хэлбэл вэб дээрх мэдээллийн учир шалтгаан хамаарлуудыг тодорхойлон ойлгуулахад ашиглагддаг бөгөөд RDF болон RDFs, XML-ээс илүү баялаг үгийн сан,



Зураг 3. Семантик вэбийн бүтэц  
Эх сурвалж: Sebastian Ryszard Kruk- Semantic Digital Libraries

семантик хэрэгслүүдээр хангаж чадах хэл юм. RDF болон OWL нь хоёулаа XML бичиглэлтэй бөгөөд OWL семантик нь RDF семантикийн өргөтгөл юм (Зураг 3-м). Хэдийгээр RDF нь XML төрлийн бичиглэлтэй боловч зориулалт болон утга нь өөр байдаг. [5]

### B. FOAFRealm (Friend-of-a-friend)

Вэб2.0 технологийн гол шийдлүүдийн нэг нь олон нийтийн сүлжээнд найзуудын холбоо хамаарлыг тогтоож хэн хэнийгээ хэр зэрэг мэдэх вэ гэдгийг тодорхойлж үнэлэх арга юм. FOAF нь RDF-д үндэслэн дүрслэгдсэн схем бөгөөд бодит амьдрал дээрх найзын холбоо хамаарлыг технологид буулгаж өгсөн хувилбар гэж ойлгож болно. Бид олон нийтийн сүлжээг ашиглаж байх үед маш олон жил уулзалдаагүй

<sup>1</sup> Семантик цахим номын сангийн төслийн хүрээнд хэд хэдэн их дээд сургуулиуд хамтран хөгжүүлж байгаа систем юм.

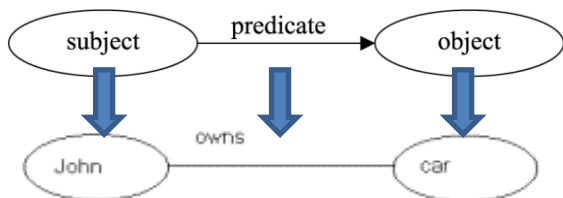
найзууд тааралддаг. Энэ нь FOAF найзууд хоорондын хамаарлын түвшинг тодорхойлон бидний хүрээллийг үүсгэн холбож өгч байгаа гэсэн үг юм.

### III. Шийдэл

Семантик цахим номын сан гэдэг нь эцсийн хэрэглэгчдийн мэдээллийн эрэл хайгуулд бодит туслалцаагаар хангах чадвартай семантик технологи бүхий цахим номын сангийн систем юм. [3]

Семантик цахим номын сан нь

- Зөвхөн цахим номын сан бус бусад системүүдтэй метадата болон харилцааны түвшинд хамтран ажиллах боломжийг хангах
- Семантик технологиор дамжуулан илүү чанартай, хэрэглэхэд хялбар, уян хатан хайлт



Зураг 2. RDF схем Эх сурвалж: W3C

хийх болон үр дүнг харуулах интерфэйс

- Олон төрлийн метадатад суурилсан мэдээллүүдийг нэгтгэх боломжтой шийдэл болон хэрэгжих юм. [4]

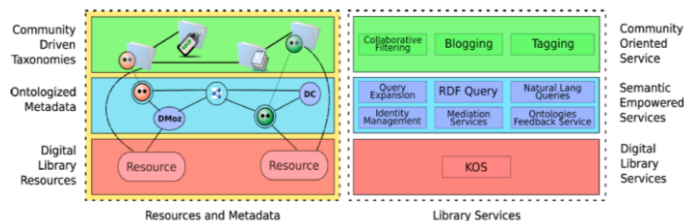
Номзүйн эдгээр үндсэн метадата тодорхойлолтуудыг семантик тодорхойлолт болгон хөрвүүлэх алхмууд нь:

1. Үндсэн номзүйн бичилтийн (Marc21) бинари файлыг задалж XML файл үүсгэнэ.
2. XSLT (*EX*Extensible *ST*ylesheet *L*anguage *T*ransformation) ашиглан үүсгэсэн XML файлыг RDF файл болгон хөрвүүлнэ
3. RDF схемийг MarcOnt семантик тодорхойлолтруу шилжүүлнэ. [4]

Хамгийн сүүлийн алхам хамгийн их ажиллагаатай бөгөөд RDF схем нэг бүрийг өөр үгийн сан ашиглан өөр нэг рүү нь хөрвүүлэх шаардлагатай болно. Өгөгдлийн эх үүсвэрүүдийн семантик хамаарлыг OWL ашиглан загварчлах бөгөөд RDF-ээр хэлбэршүүлсэн өгөгдөлтэй ажиллахад SPARQL хэлийг ашиглана.

Цахим номын сангийн системд 3 түвшний метадата бүтцийг Зур.4-т дүрслэн харуулсан байна. Давхарга бүр үндсэн семантик мэдээллүүдийг агуулах бөгөөд хайлт хийх болон үр дүнг харуулах нэмэлт чадавхыг хангах чадвартай байх юм.

Хамгийн доод давхарга нь бодит нөөц, тэдгээрийн бүтэц болон эх сурвалжуудыг хадгалж байгаа бүтцийн онтологийг хэсэг. Дунд давхарга нь албан ёсны номзүйн бичилтийг семантик түвшинд тодорхойлох нэмэлт онтологийн давхарга бөгөөд баримтын метадатаг удирдах, хүлээн авах хадгалах үйлчилгээгээр хангана. Хамгийн дээд давхарга нь хэрэглэгчдийн бүлгэмийн нөөцүүдийг үнэлэх, тэмдэглэх үйл ажиллагаанд суурилан мэдээллийг дундаа хуваалцах, мэдлэгийг бий болгоход уншигчдийн хувь нэмрийг оруулахад чиглэсэн бүлгэмийн онтологийн давхарга юм.



Зураг 4. Семантик цахим номын сангийн метадата болон үйлчилгээний архитектур

JeromeDL системийн доорх үндсэн бүрэлдэхүүн хэсгүүдийг агуулдаг. Үүнд:

- Нөөцийн удирдлага:** Цахим санд нөөцийн метадата болон агуулга хоёуланг нь хадгална. JeromeDL системийн хувьд үндсэн онтологийн дагуу семантик тодорхойлолтуудаар нөөц бүрийг дүрсэлдэг. Шинээр нэмэгдсэн нөөц бүрийг шалгах бөгөөд энэ нь тухайн системийн агуулгын чанарыг баталгаажуулж байгаа юм.
- Хайлтын боломж:** JeromeDL нь семантик вэб өгөгдөлд суурилсан хайлтын болон үр дүнг харуулах боломжуудаар хангагддаг. Энэхүү системийн хайлтын алгоритм нь стандарт болон нэмэлт хайлт гэсэн хэсгээс бүрдэх бөгөөд 3 үндсэн алхмыг агуулдаг бөгөөд алхам бүрт өөр өөр метадата эх сурвалжууд шаардлагатай болдог.
  - Бүтэн бичвэрийн хайлт хийх
  - Номзүйн бичилтийн хайлт хийх
  - Хэрэглэгчид чиглэсэн семантик хайлт

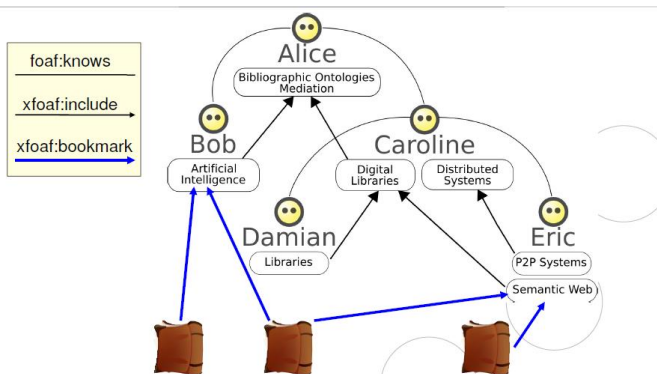
```

procedure SEMANTIC SEARCH ( File RDFfile , String query)
model.read(RDFfile, null);
queryString = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> "
+ "PREFIX dcterns:<http://purl.org/dc/terms/> "
+ "SELECT ?title ?type"
+ " WHERE { ?person dcterns:title ?title . ?person
rdf:type ?type. ";
+ " { ?person dcterns:<criteria> ? <criteria> FILTER
regex(?<criteria>, " <query>".i) } }";
Query query1 = QueryFactory.create(queryString);
QueryExecution qe = QueryExecutionFactory.create(query1,
model);
ResultSet results = qe.execSelect();
End procedure SEMANTIC SEARCH

```

Зураг 5. Семантик хайлтын алгоритм  
Эх сурвалж: Mohamed Kholif, Suaad Awad Mohamed, A Semantic Assistant Agent for Digital Libraries

С. *Хэрэглэгчийн мэдээллийн удирдлага буюу Семантик олон нийтийн хамтын ажиллагаа*<sup>2</sup>: Нөөцүүдийг нэмэлт семантик тодорхойлолтуудаар хангахын тулд FOAF-д суурилсан хэрэглэгчийн удирдлагыг ашигладаг. Хэрэглэгчид хадгалагдсан нөөцүүдийг үнэлэх, төрөлжүүлэх, тэмдэглэх боломжтой юм. Хэрэглэгчийн мэдээллийн түүх болон түүний найзуудын мэдээллүүд автоматаар цуглуулагдан боловсруулагдах бөгөөд эргээд тухайн хэрэглэгчид ямар нөөцийг санал болгоход ашиглагдах юм. Нөөц бүрийг ямар нэг төрөлд харъяалагдах бөгөөд цуглуулсан нөөцийн төрөлд үндэслэн уншигчийн сонирхлыг тодорхойлно. JeromeDL системийн хувьд өөрийн найзуудын тэмдэглэгээг харах, зарим төрлүүдийг өөрийн бүтэцтэй холбох боломжтой юм. Дараа нь уншигчийн үүсгэсэн төрөл бүрт давтагдашгүй ACL (*Access control list*) оноох бөгөөд энэ нь тухайн төрлийн агуулгыг аль найз нь харах эсвэл харах боломжтойг тодорхойлох юм. ACL-н оролтууд нь тухайн төрлийг үүсгэгч болон уншигчийн хоорондын холбоо болон төсөөтэй байдлын түвшинд үндэслэнэ. [4]



Зураг 6. Семантик олон нийтийн хамтын ажиллагааны жишээ

D. *Харилцааны холбоос*: Цахим номын сангийн сүлжээгээр хайлт хийх замаар гадаад орчинтой холбоо тогтоох боломжтой. JeromeDL-н өгөгдлийн сангийн агуулга нь зөвхөн цахим номын сангийн вэб хуудаснуудаас гадна бусад цахим сангийн вэб аппликэшнүүдээс ч хайлт хийх боломж бүхий ELP (*Extensible library protocol*) тусгай вэб үйлчилгээний интерфэйсийг хөгжүүлж байна. ELP гол санаа нь цахим номын сангийн олон төрлийн орчинг холбох боломжтой болгох явдал юм. Номын сан нэг бүр өөр нэг номын сангийн агуулгыг

мэддэг байх хэрэгтэй бөгөөд энэ нь L2L (*Library to Library*) буюу сүлжээний P2P сүлжээний нэгэн адил бүтцийг үүсгэх юм. Хамгийн гол нь эдгээр системүүд нь дор хаяж DublinCore метадатаг дэмждэг байх хэрэгтэй юм. Хэрвээ холбогдож байгаа цахим номын сангийн системүүд нь нөөцөө семантикаар тодорхойлдог байвал тэдгээрийн хоорондын холбоос нь автоматаар семантик тодорхойлолтын түвшинд хийгдэнэ. Энэ нь тухайн L2L сүлжээгээр дамжуулан семантик хайлтын алгоритмуудыг ашиглах боломжтой болгох юм.[4]

#### IV. ДҮГНЭЛТ

Судалгааны ажлын хүрээнд вэбийн нөөцийн зохион байгуулалтыг илүү боловсронгуй түвшинд хүргэх семантик технологийг цахим номын сангийн системд ашиглах талаарх онолын судалгаануудыг авч үзсэн болно. Цахим номын сангийн хувьсалд онтологи хэмээх ойлголт хамгийн гол үүргийг гүйцэтгэж байгаа бөгөөд машин хүнтэй адил мэдээллийг боловсруулж, ойлгож хариу өгөх боломжтой болох юм.

Онолын хүрээнд хийгдсэн судалгааны ажил нь цахим нөөцийг илүү бүтээлчээр ашиглах болон олон нийтийн оролцоотойгоор бодит мэдлэгийг бий болгох орчинг бүрдүүлэхэд семантик технологи болон цахим номын сангийн системүүдийг хэрхэн хослуулах болохыг тодорхойлсон. Энэхүү шийдэлтэй холбоотойгоор олон төрлийн метадатаг нэгтгэн хөрвүүлэх болон өөр өөр төрлийн цахим сангуудын хамтын ажиллагааг хангахад чиглэсэн судалгаа шинжилгээ хийгдсээр байна. Судалгааны ажлын цаашдын зорилт нь их дээд сургуулиудын номын сангийн системүүд болон эрдэм шинжилгээний ажлуудын санг судлах, ашиглаж буй системүүд болон тэдгээрийг холбох нэгтгэх шийдлийг тодорхойлохоос гадна семантик технологийн онтологи түвшингүүдийг нарийвчлан тодорхой судлах явдал юм.

#### НОМЗҮЙ

- [1] M. Barbera, M. Nucci, D. Hahn, and C. Morbidon, A Semantic Web Powered Distributed Digital Library system, 2008.
- [2] М.Цэндсүрэн, Таны өмнөөс семантик вэб, 2009
- [3] P. Rajiv, Dr. D. Sanjav, Interoperability between semantic web layers: Acommunicating agent approach, nternational journal of computer application, Vol. 12-no.3, pp.28-32, November 2010.
- [4] Sebastian Ryszard Kruk, Stefan Decker, and Lech Zieborak, JeromeDL - Adding Semantic Web Technologies to Digital Libraries
- [5] Peter F. Patel-Schneider, What is OWL (and why should I give a hoot)?, 2005
- [6] Ж.Цэзэн, Монголын номын сангийн өнөөгийн байдал, 2013

<sup>2</sup> Semantic social collaborative filtering