An Open Localisation Interface to CMS using OASIS Content Management Interoperability Services

Aonghus Ó hAirt, Dominic Jones, Leroy Finn and David Lewis

Centre for Next Generation Localisation, TCD

Introduction

- Content Management Systems (CMS) are used by a wide range of organisations to maintain web sites and other sources of content/documentation
- Organisations may maintain several CMS for different departments and in different locations
  - e.g. marketing, user documentation, locale-specific
- Maintaining and localising content across CMS becomes increasingly important as enterprises seek a better overall content experience of their consumer

Background: Content Management Systems

- CMS facilitate the creation, storage, editing, and publishing of content
- Web Content Management Systems (WCMS) are typically used for web sites and blogs
  - e.g. Drupal, WordPress, Joomla
- Enterprise Content Management Systems (ECM) are used to manage and store content related to organisations' processes
  - e.g. Alfresco, Sharepoint, Nuxeo

Background: Content Management

- CMS can vary in many ways:
  - Platform/Languages: Java, PHP, ASP.NET, Perl, Python, Ruby on Rails etc.
  - Licence: commercial, open source
  - Cost
  - Performance
  - Scalability
  - Functionality: depth and breadth of feature set
  - (Technical) Support: professional customer service (commercial) vs. community support (open source)
  - Systems support: integration with other tools and technologies

Background: Content Management

- The strengths and weaknesses of CMS for particular purposes can lead to multiple CMS being deployed in separate areas of organisations for different purposes.

Content Management & Localisation

- Content to be Localised increasingly sourced and then published via CMS
- Good integration of CMS with Localisation Tools can reduce overall localisation costs
- Increasingly need to localise content that is still under revision - requires asynchronous content status tracking between CMS and Localisation Tools

CMS Interoperability

- Integrating with CMS requires the use of an API. Until now, most CMS used proprietary APIs
- Proprietary interfaces to CMS lead to limited support, vendor lock-in and poor interoperability between CMS and with localisation tools
- Content Management Interoperability Service (CMIS) from OASIS offers a standardised API for interacting with CMS
- Localisation is excluded the scope of CMIS

How can CMIS facilitate the localisation of content across multiple CMS?

OASIS Content Management Interoperability Services (CMIS)

- “defines a domain model and Web Services and Restful AtomPub bindings that can be used by applications to work with one or more Content Management repositories/systems.” (CMIS standard)
- Published in 2010
- Participation from Adobe, Alfresco, EMC, IBM, Microsoft, Oracle, SAP, and others.

CMIS Implementations (server support)

Alfresco
Apache Chemistry InMemory Server
Athento
COI
Day Software CRX
EMC Documentum
eXo Platform with xCMIS
Fabasoft
HP Autonomy Interwoven Worksite
IBM Content Manager
IBM FileNet Content Manager
IBM Content Manager On Demand
IBM Connections Files
IBM LotusLive Files
IBM Lotus Quickr Lists
ISIS Papyrus Objects
KnowledgeTree
Maarch
Magnolia (CMS)
Microsoft SharePoint Server 2010
NCMIS
NemakiWare
Nuxeo Platform
O3spaces
OpenIMS
OpenWGA
PTC Windchill
SAP NetWeaver Cloud Document
Seapine Surround SCM 2011.1
Sense/Net
TYPO3
VB.CMIS

CMIS Objects

- A repository is a container of objects.
- Objects have four base types:
  - **Document object** – “elementary information entities managed by the repository”
  - **Folder object** – “serves as the anchor for a collection of *file-able* objects”
  - **Relationship object** – “instantiates an explicit, binary, directional, non-invasive, and typed relationship between a *Source Object* and a *Target Object”*
  - **Policy object** – “represents an administrative policy that can be enforced by a repository, such as a retention management policy.”

(CMIS Specification)

CMS-L10n Interoperability: Two Requirements

- Internationalisation meta-data allows content authors to specify instructions to inform L10n processes
- W3C Internationalization Tag Set (ITS) provides standard content mark-up rules
  - Internal and external rules
- Aim to support external ITS rules via CMIS

- Need to signal L10n-relevant updates to documents
- ITS successor workgroup identified a requirement for such ‘readiness’ signalling
- Aim to support open asynchronous change notification for CMIS

Handling ITS Rules in CMS

- ITS uses Xpath selectors to indicate elements within a document subject to specific localisation instructions
- Need to capture document-level ITS rules
  - e.g. <its:translateRule translate="no" selector="/text/head"/>
- External document-level rules can be associated with files "by associating the rules and the document through a tool-specific mechanism" (W3C ITS, 2007)
- With CMIS we enable:
  - The same rule to be applied to multiple documents
  - Multiple rules to be applied to individual documents
  - Specify the precedence order in which rules are processed for a document

Signalling Readiness from CMS

- Readiness meta-data
  - Indicates the readiness of a document for submission to L10n processes or provide an estimate of when it will be ready for a particular process
- Data model
  - **ready-to-process** – type of process to be performed next
  - **process-ref** – a pointer to an external set of process type definitions used for ready-to-process
  - **ready-at** – defines the time the content is ready for the process, it could be some time in the past, or some time in the future
  - **revised** – indicates is this is a different version of content that was previously marked as ready for the declared process
  - **priority –** high or low
  - **complete-by –** indicates target date-time for completing the process

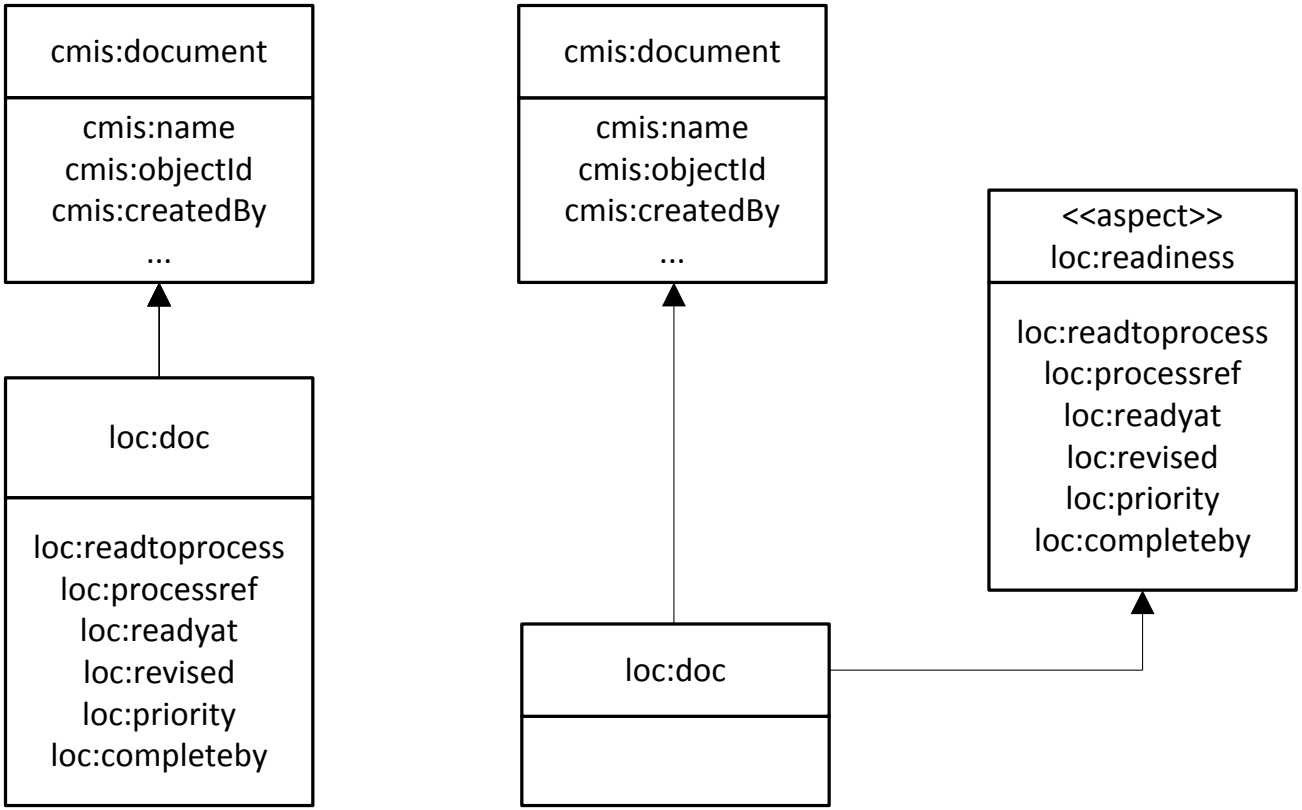
Polling extension to CMIS

- Polling schemes describe the way in which documents are polled for updated readiness properties
  - scheme name / ID
  - polling interval
  - notification method
  - notification target / host
  - port (for network connection)
  - readiness property
  - readiness value

Design: Extending CMIS Implementations

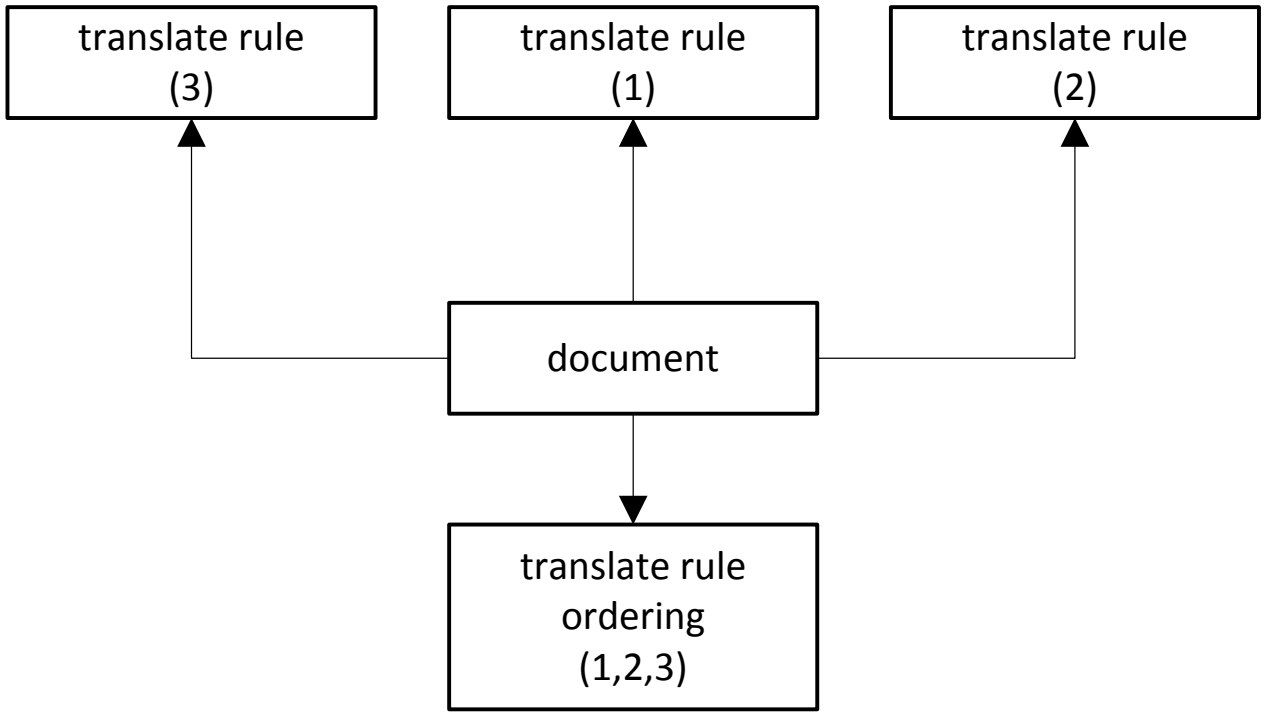
- Two approaches to modelling the localisation information:
  - Custom content modelling
  - Alfresco aspects
- Implementation in repository
  - Alfresco (primary)
  - Nuxeo (basic testing)

Readiness



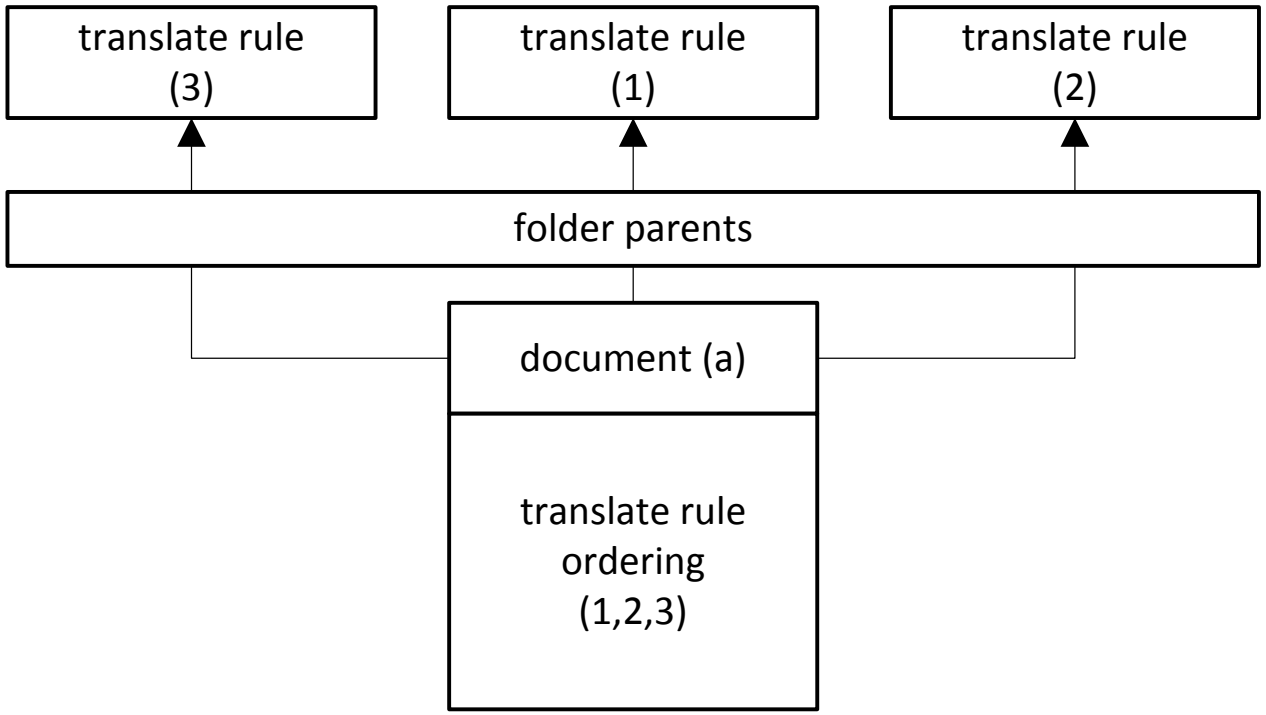
Readiness modelled as custom object (left) and with an aspect (right)

Translate rules



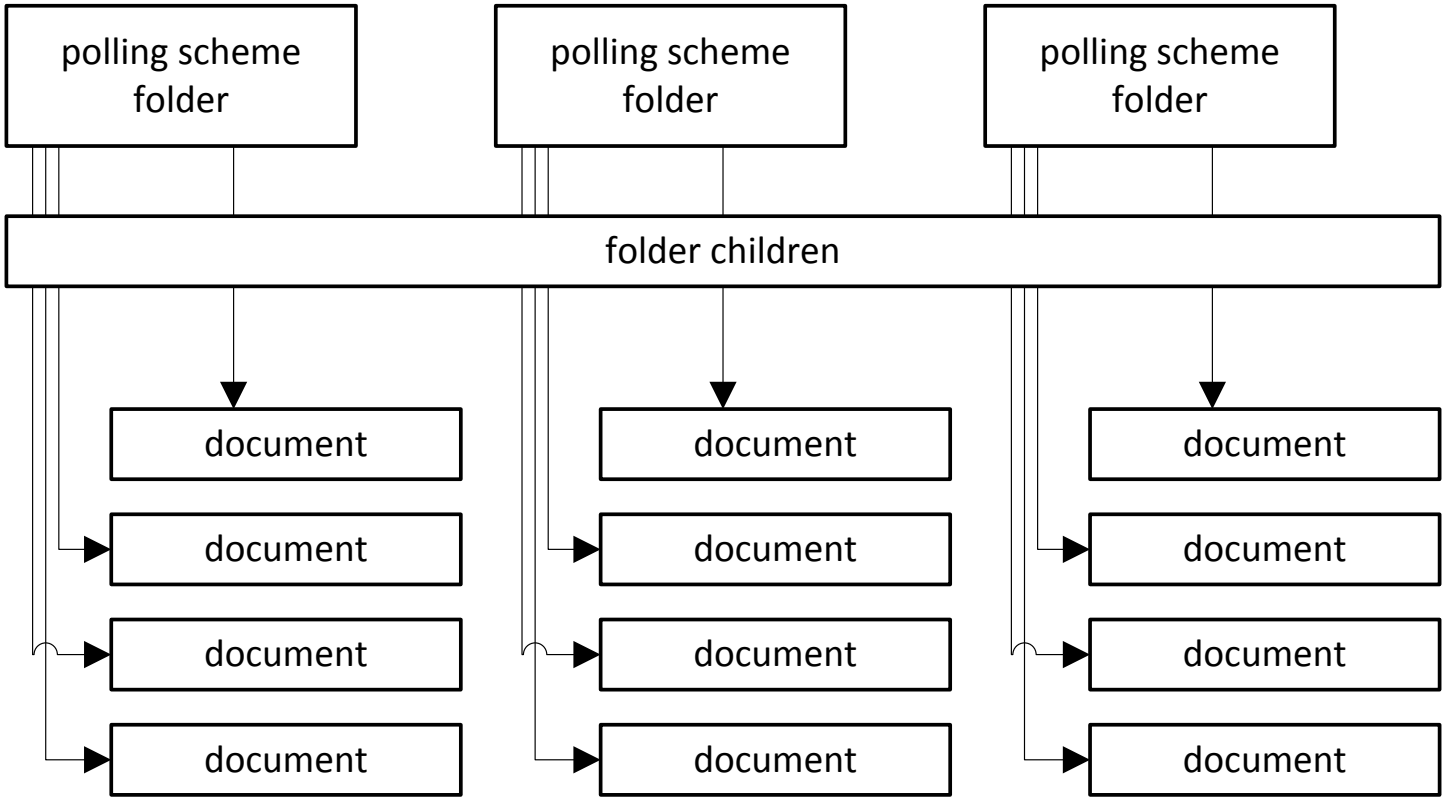
Translate rules as policy objects

Translate rules

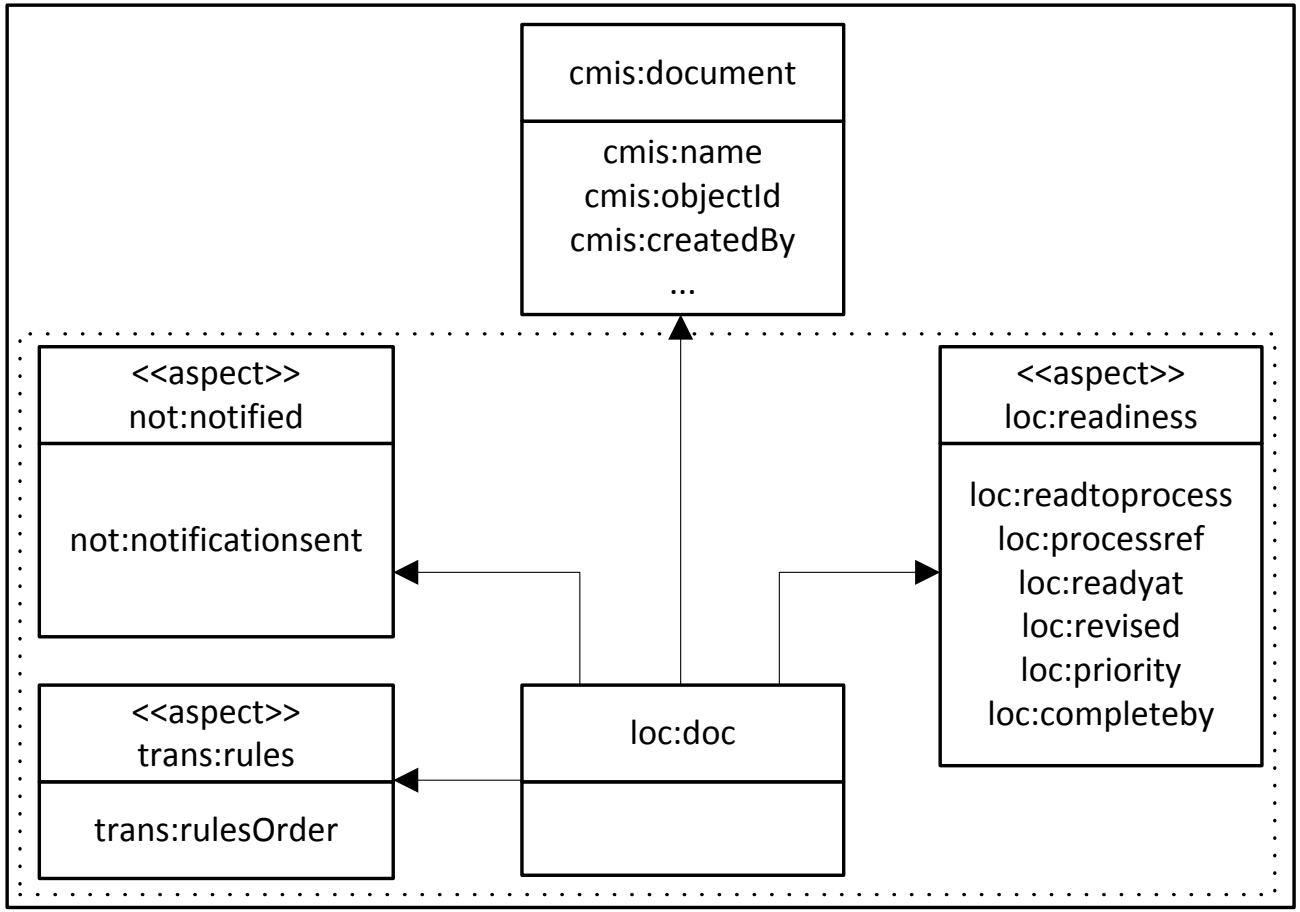


Translate rules as folder objects

Polling Schemes



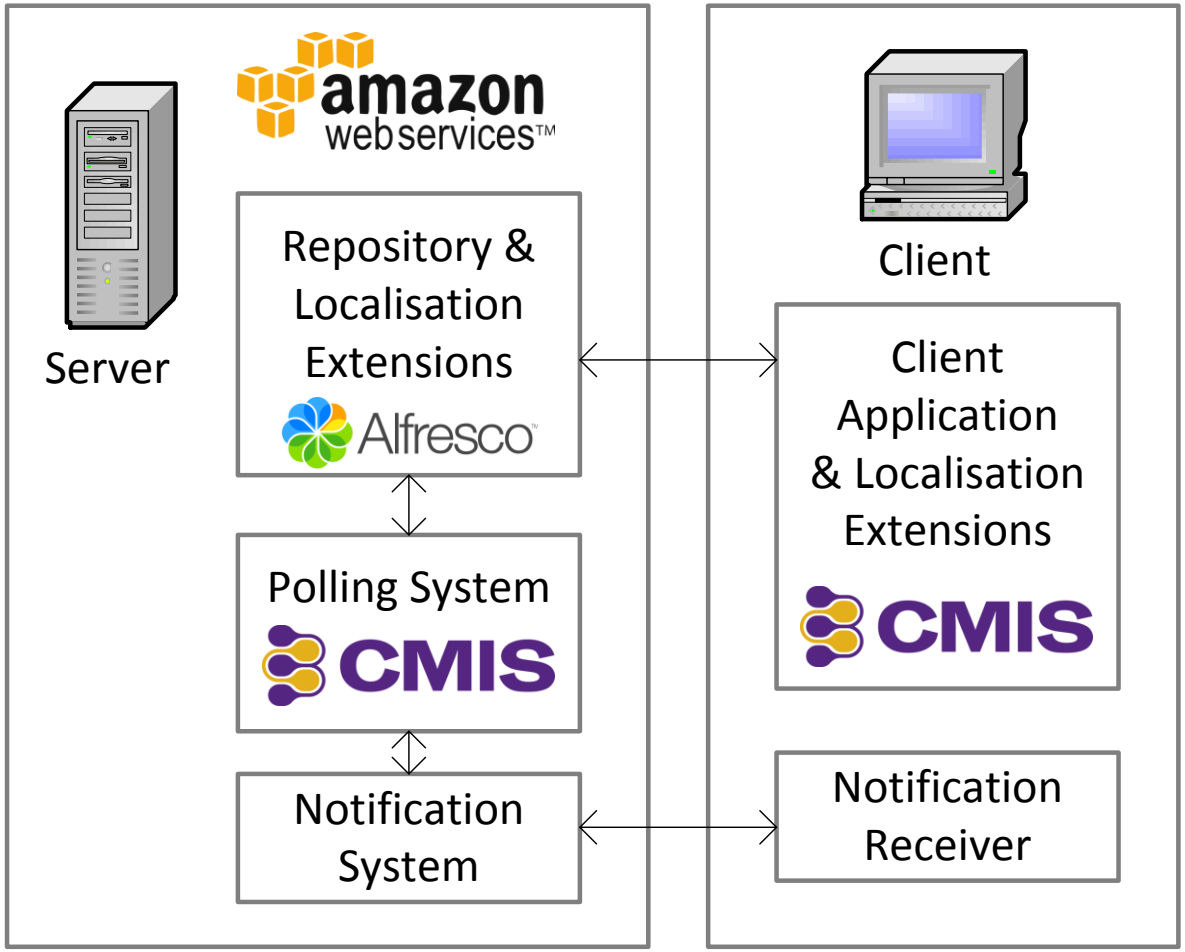
Document model with localisation



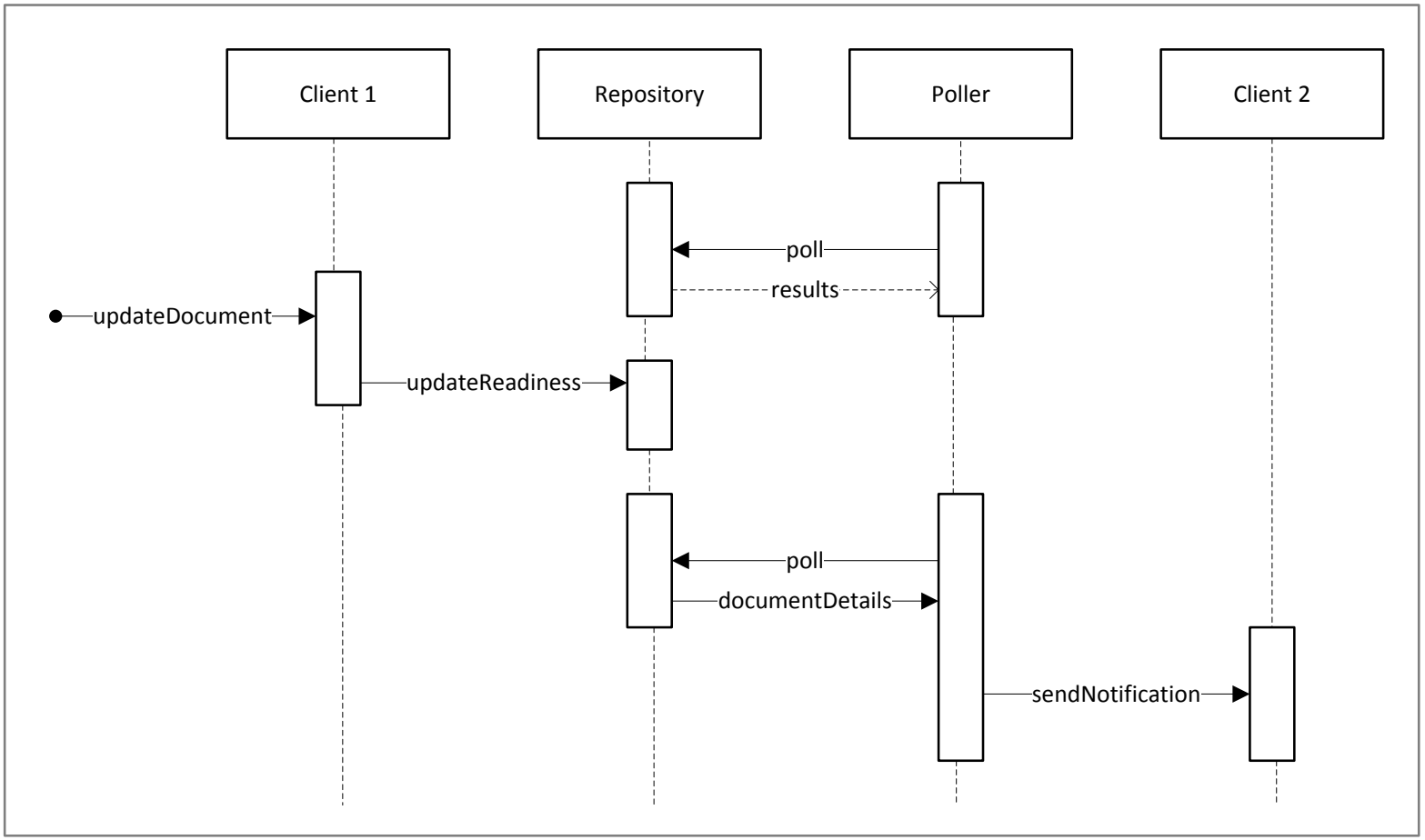
Applications

- Repository browser tool
- Polling system
- Notification system
- Test tools

Technical setup

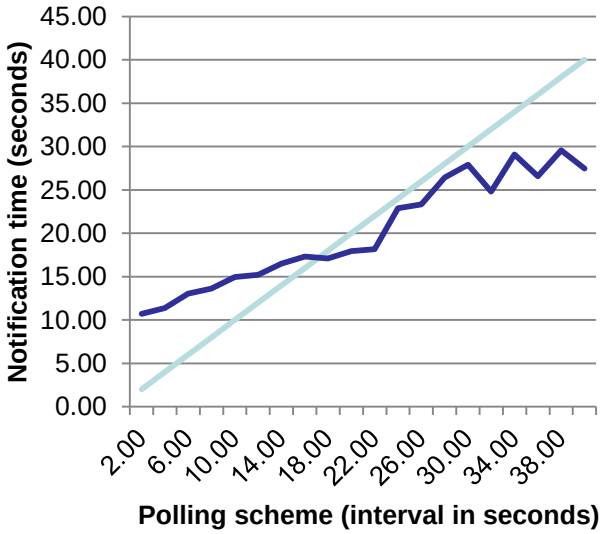


Polling sequence



Evaluation

- Notification response time



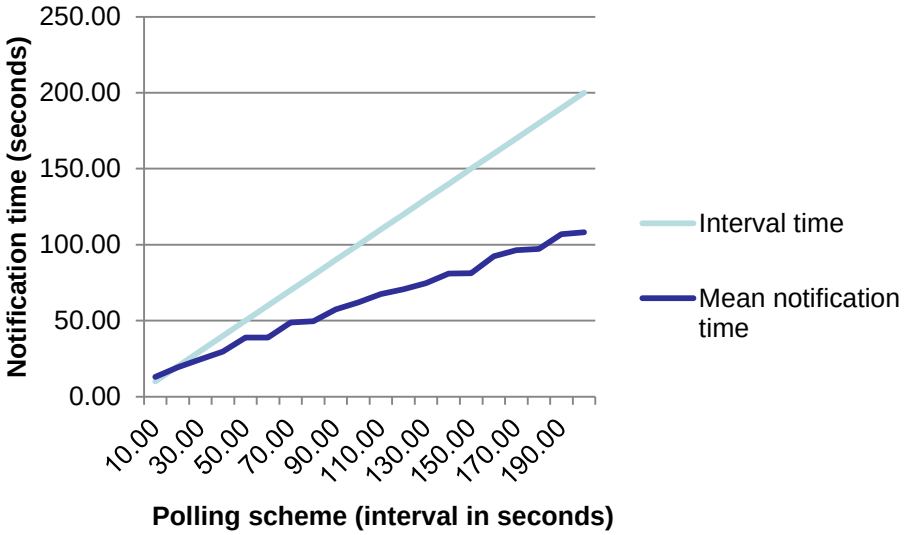


Table 5.3: Results for initial *TestSimulationTimers* experiment

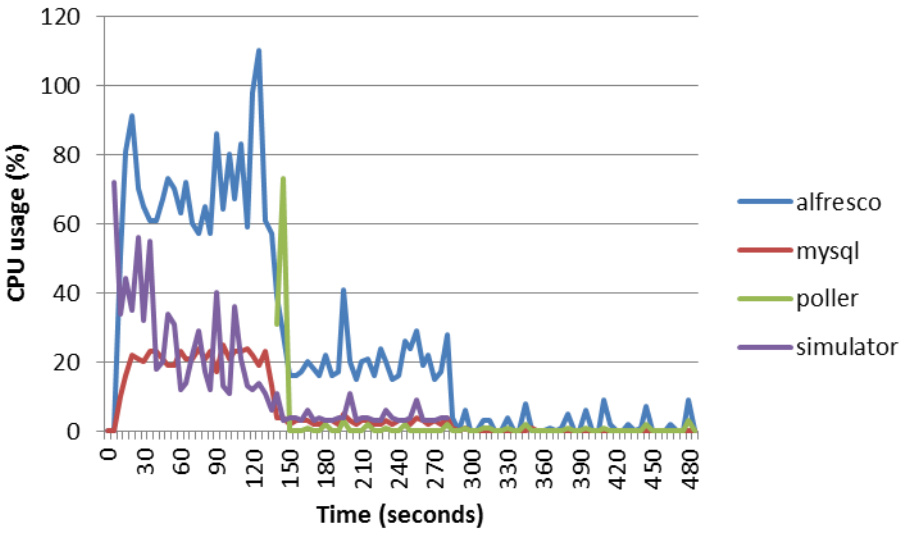
Scheme	Interval	Mean	Std.Dev.	Early		Late	
PS1	2.00	10.72	4.85	0	0%	50	100%
PS2	4.00	11.36	4.70	2	4%	48	96%
PS3	6.00	13.04	4.93	6	12%	44	88%
PS4	8.00	13.62	4.88	11	22%	39	78%
PS5	10.00	14.94	4.89	12	24%	38	76%
PS6	12.00	15.20	5.41	15	30%	35	70%
PS7	14.00	16.50	5.74	19	38%	31	62%
PS8	16.00	17.30	5.56	20	40%	30	60%
PS9	18.00	17.10	6.21	31	62%	19	38%
PS10	20.00	17.96	6.68	34	68%	16	32%
PS11	22.00	18.18	5.83	43	86%	7	14%
PS12	24.00	22.88	7.73	26	52%	24	48%
PS13	26.00	23.34	8.99	32	64%	18	36%
PS14	28.00	26.42	9.38	28	56%	22	44%
PS15	30.00	27.92	10.26	28	56%	22	44%
PS16	32.00	24.82	10.16	36	72%	14	28%
PS17	34.00	29.10	11.44	32	64%	18	36%
PS18	36.00	26.58	11.20	39	78%	11	22%
PS19	38.00	29.58	11.77	36	72%	14	28%
PS20	40.00	27.48	14.26	38	76%	12	24%

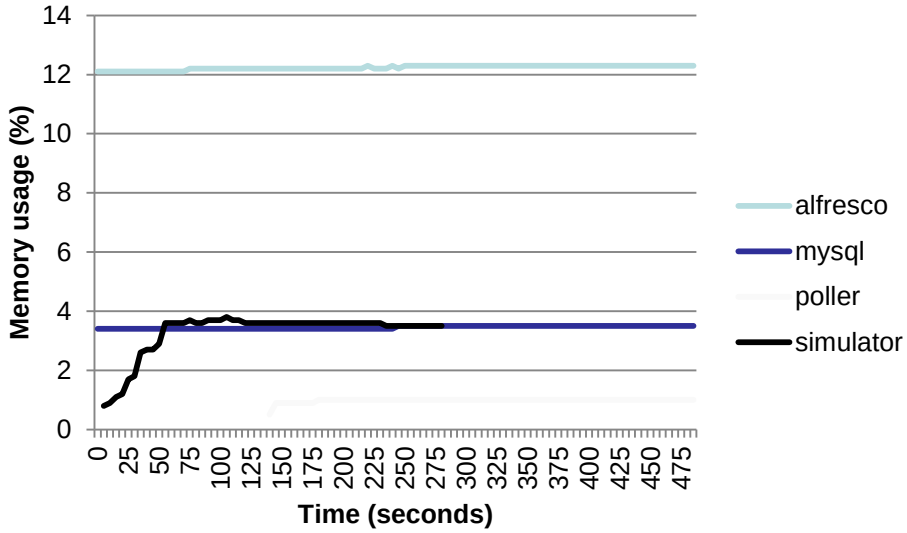
Table 5.4: Results for second *TestSimulationTimers* experiment

Scheme	Interval	Mean	Std.Dev.	Early		Late	
PS1	10.00	13.06	5.18	17	34%	33	66%
PS2	20.00	19.22	6.89	31	62%	19	38%
PS3	30.00	24.52	9.75	35	70%	15	30%
PS4	40.00	29.56	12.62	39	78%	11	22%
PS5	50.00	38.86	14.83	36	72%	14	28%
PS6	60.00	38.84	18.37	44	88%	6	12%
PS7	70.00	48.70	20.50	39	78%	11	22%
PS8	80.00	49.58	23.75	44	88%	6	12%
PS9	90.00	57.36	26.01	43	86%	7	14%
PS10	100.00	62.06	29.17	43	86%	7	14%
PS11	110.00	67.56	31.46	45	90%	5	10%
PS12	120.00	70.78	34.60	46	92%	4	8%
PS13	130.00	74.68	38.89	46	92%	4	8%
PS14	140.00	80.92	40.38	46	92%	4	8%
PS15	150.00	81.26	43.91	50	100%	0	0%
PS16	160.00	92.34	45.88	47	94%	3	6%
PS17	170.00	96.32	50.41	47	94%	3	6%
PS18	180.00	97.26	51.99	49	98%	1	2%
PS19	190.00	106.84	54.22	47	94%	3	6%
PS20	200.00	108.20	58.42	49	98%	1	2%

Evaluation

- Performance evaluation





Conclusion

- Have extended CMIS to support:
  - Document level ITS rules
  - Open document change notification mechanism
- Strong potential to streamline CMS-L10n integration
- Achieved with current CMIS specification
  - Custom extension to folder object
  - Custom extension to policy object may be better
- Optimised implementation using Alfresco aspects
- Plans:
  - Broader support for ITS rule types
  - Integrate with XLIFF
  - Discuss extensions with CMIS-compliant vendors

Questions.

THANK YOU.

Follow ITS Use Case at:

http://www.w3.org/International/multilingualweb/lt/wiki/CMS_Neutral_External_ITS_Rules_and_Readiness