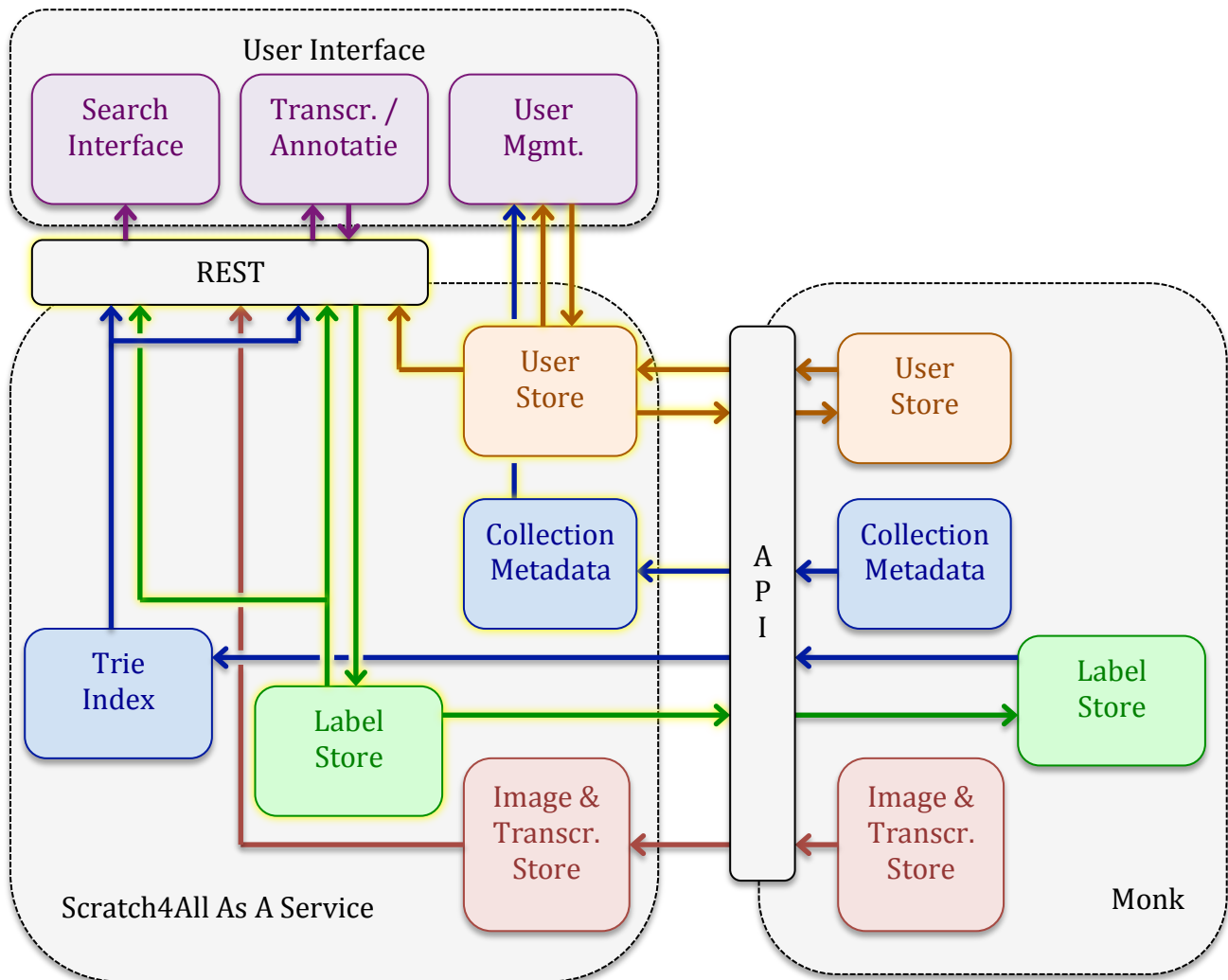


Architectuur Scratch4All As A Service – 0.1

Rolf Fokkens, Target Holding

20 juli 2012

Overzicht



Bovenstaande diagram geeft de architectuur weer van Scratch4All As A Service (S4AAAS). De gele highlights markeren delen die nog moeten worden ontwikkeld.

S4AAAS ondersteunt een aantal User Interface schermen die het resultaat zijn van het Scratch4all project:

- De Search Interface
- De Transcriptie/Annotatie Interface
- Het User management scherm

Deze Search Interface en de Transcriptie/Annotatie schermen zijn aan S4AAAS gekoppeld via een REST interface, analoog aan de architectuur zoals opgeleverd in Scratch4All.

S4AAAS bevat 4 “data stores”:

- **User Store**
De User Store is een duplicaat van de Monk User Store met authenticatie en autorisatie gegevens, die noodzakelijk is om S4AAAS autonoom te kunnen laten werken. Deze User Store synchroniseert met de Monk User Store.
Deze User Store moet nog worden gebouwd (zie de gele highlight)
- **Collection Metadata**
De Collection Metadata is een duplicaat van de Monk Collection Metadata. De Collection Metadata is nodig ter ondersteuning van de User Store, i.h.b. de autorisatie.
- **Trie Index**
Dit is de Search Trie die op dit moment wordt gebruikt voor de bestaande Search Interfaces. De Trie Index wordt 's nachts opgebouwd op basis van de Monk Label Store.
De Trie Index ontwikkeld binnen het Scratch4All project.
- **Label Store**
Deze Label store is een tijdelijke buffer die dient om wordlabels binnen S4AAAS op te slaan zolang Monk ze nog niet heeft verwerkt. De labels in de label store worden uitgewisseld met Label Store van Monk, zodra de labels in de Trie Index staan worden ze uit de Label Store verwijderd.
Deze Label Store moet nog worden gebouwd (zie gele highlight)
- **Image & Transcription Store**
S4AAAS gebruikt eigen kopieën nodig van Images en Line- en Pagetranscriptions, zoals die ook beschikbaar zijn in de Monk “Image & Transcription Store”. S4AAAS heeft echter voldoende aan lagere kwaliteit images dan die van Monk.

Afwijkingen t.o.v. Scratch4All

Uitgezonderd de geel gemarkeerde onderdelen is er nog een aantal verschillen t.o.v. Scratch4All:

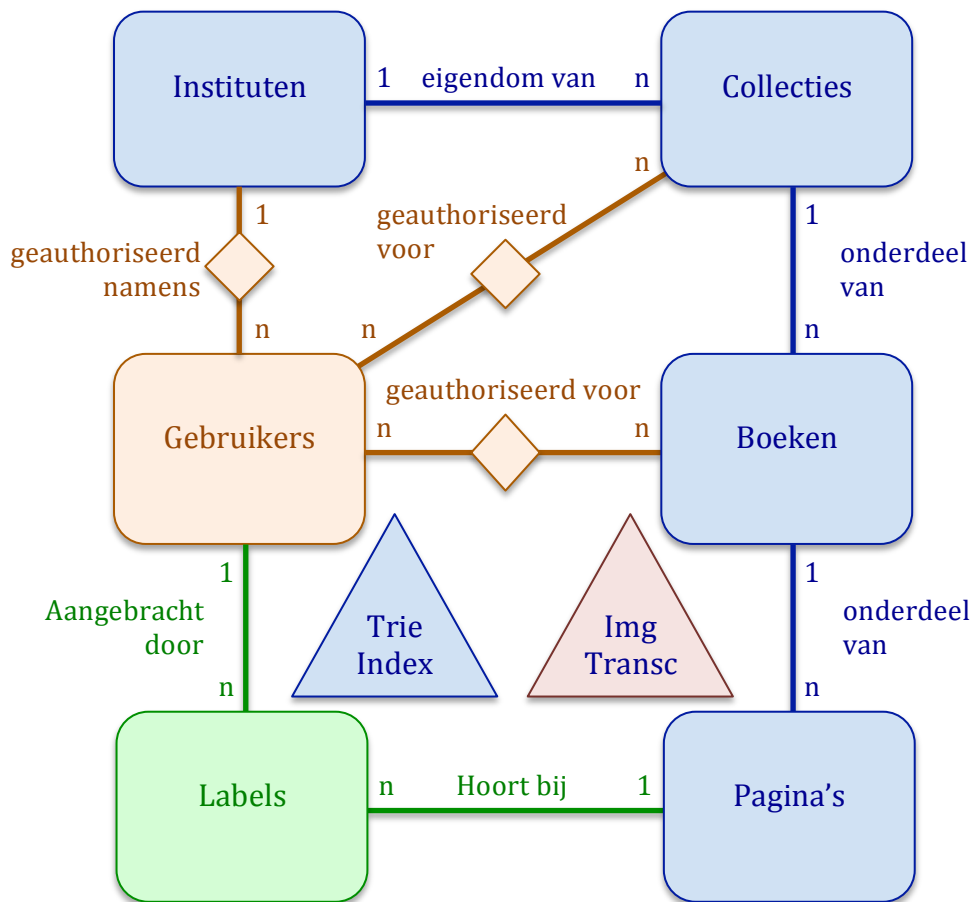
- De S4AAAS Label Store maakt het mogelijk vers ingevoerde labels onmiddellijk via de Search Interface op te leveren als zoek resultaat.
- De S4AAAS Label Store maakt het mogelijk om (de workflow van) Expert reviews van user Labels eenvoudig mogelijk te maken.
- De Image & Transcription Store in Scratch4All dient als een cache, en wordt dus alleen gevuld op basis van zoek opdrachten. In S4AAAS moet de Image & Transcription Store op voorhand worden gevuld.

Detail beschrijvingen datastores

De datastores van S4AAS worden hier in nader detail beschreven. Dit wordt (soms) gedaan in termen van tabellen en datamodellen, maar de feitelijke implementatie staat open. Ter vergelijking: Monk zelf werkt prima op basis van data in files.

Datamodel

Het datamodel voor S4AAS ziet er als volgt uit:



In dit model zijn de de genoemde datastores verenigd, en onderling met elkaar in relatie gebracht.

De Trie Index en de Image Transcription Store hebben een bijzondere positie in het datamodel, omdat zij middels bestanden worden opgeslagen.

Algemene overwegingen t.a.v. uitwisseling met Monk

De details worden verder op uitgewerkt. Ten aanzien van de onderdelen "Gebruiker" en "Labels" zijn er echter aanvullende overwegingen: In het datamodel is expliciet zichtbaar wat de status is van de uitwisseling met Monk.

- Alle Gebruikers en Labels worden voorzien van een tijdstempel van de laatste wijziging aan de zijde van S4AAAS. Dit tijdstempel impliceert dat er moet worden gesynchroniseerd.
- Verwijderde Gebruikers aan de zijde van S4AAAS wordt expliciet (met een kolom 'Deleted') gemarkeerd, maar nog niet echt verwijderd.
- De informatie in Monk wordt tijdens de synchronisatiemomenten vergeleken met de Gebruikers en Labels (via de Trie Index) binnen S4AAAS. Voor S4AAAS-wijzigingen die door Monk zijn verwerkt, worden de tijdstempels gewist, Verwijderde gebruikers worden daadwerkelijk aan de zijde van S4AAAS verwijderd.
- Labels die binnen Monk zijn opgeslagen, worden verwijderd uit S4AAAS (omdat ze in de Trie Index beschikbaar zijn).

Details

Instituten	id	NUMBER	
	monk_id	VARCHAR2	e.g. "NA"
	short_name	VARCHAR2	e.g. "NA"
	long_name	VARCHAR2	e.g. "Nationaal Archief"
Collecties	id	NUMBER	
	instituut_id	NUMBER	REFERENCES Instituten (id)
	monk_id	VARCHAR2	e.g. "KdK"
	long_name	VARCHAR2	e.g. "Kabinet der Koningin"
Boeken	id	NUMBER	
	collectie_id	NUMBER	REFERENCES Collecties (id)
	monk_id	VARCHAR2	e.g. "NL-HaNA_2.02.04_3990"
	long_name	VARCHAR2	e.g. "Index op Koninklijke besluiten - 1893..."
	handle_url sheer	VARCHAR2 NUMBER	e.g. "http://proxy.handle.net/..."
Paginas	id	NUMBER	
	boek_id	NUMBER	REFERENCES Boeken (id)
Gebruikers	id		
	monk_id	VARCHAR2	e.g. "rfokkens"
	password	VARCHAR2	Encrypted
	permissions	NUMBER	global_permissions
	byuser_id	NUMBER	REFERENCES users (id)
	timestamp deleted	DATETIME VARCHAR2	
GebrInst	gebruiker_id	NUMBER	REFERENCES Gebruikers (id)
	instituut_id	NUMBER	REFERENCES Instituten (id)
	permissions	NUMBER	institution_permissions
	byuser_id	NUMBER	REFERENCES users (id)
	timestamp	DATETIME	
	deleted	VARCHAR2	
GebrColl	gebruiker_id	NUMBER	REFERENCES Gebruikers (id)
	collectie_id	NUMBER	REFERENCES Collecties (id)
	permissions	NUMBER	collection_permissions
	byuser_id	NUMBER	REFERENCES users (id)
	timestamp	DATETIME	
	deleted	VARCHAR2	
GebrBoek	gebruiker_id	NUMBER	REFERENCES Gebruikers (id)
	boek_id	NUMBER	REFERENCES boeken (id)
	permissions	NUMBER	book_permissions
	page_from	NUMBER	REFERENCES paginas
	page_to	NUMBER	REFERENCES paginas
	byuser_id	NUMBER	REFERENCES users (id)
	timestamp	DATETIME	
	deleted	VARCHAR2	

Labels	id	NUMBER	
	pagina_id	NUMBER	REFERENCES paginas (id)
	byuser_id	NUMBER	REFERENCES users (id)
	timestamp	DATETIME	
	x1	NUMBER	
	x2	NUMBER	
	y1	NUMBER	
	y2	NUMBER	
	label_text	VARCHAR2	