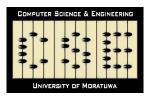


The Common Protocol for Document Verification built on Digital Signatures

http://www.idstack.one







The Common Protocol for Document Verification built on Digital Signatures

Our Team









Supervisors





Prof. Gihan Dias

PhD (UCD), MSc (UCSB), BSc Eng. (Hons) (Moratuwa), MIE (SL), CEng Eng. Dr. Shantha Fernando

PhD (TU Delft), MPhil (Moratuwa), BSc Eng. (Hons) (Moratuwa), MIE (SL), MIEE (UK), CEng

Chanaka Lakmal

Sachithra Dangalla

Chamin Wickramarathna

Chandu Herath

id tack...



Problem



In countries like Sri Lanka the document verification process is highly based on the printed copies.

Problem: 1/4

Documents have unpredictable layouts that have raw information and there is no mechanism in the current research field to extract them into common format of machine readable.



Problem: 2/4

The document is subjected to a process of verification by an authority; using a hand written signature.

Another verified copy? The whole process is repeated.



Non existence of a mechanism in the current research field that can calculate trustworthiness of a individual document.

Sometimes the fraudulent activities that are followed by certain citizens leading to doubtfulness of the authenticity of the content.





Non existence of a mechanism in the current research field that can calculate correlation among set of documents belonging to a person.



Revolutionize

Document Processing



Features



idstack

Extraction

Extract and create structured document

Validation

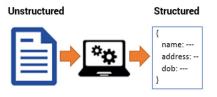
Validate document using digital signatures

Confidence score

Measure trustworthiness of a individual document

Correlation score

Measure correlation among set of documents







Extraction

Provide an application to extract data

- from a paper document
- from a digital document (PDF)

and convert into a Machine Readable (.idstack) format and verify it.



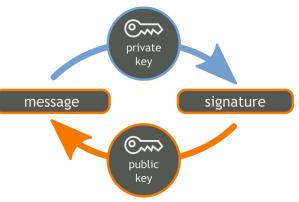
tock



idstack

Validation

Provide an application to validate and verify a document with the use of digital signatures.







Confidence Score

Provide an application with algorithm to

calculate the trustworthiness of a individual document







Correlation Score

Provide an application with algorithm to

calculate the correlation among a set of documents







All the citizens are grouped into one of the followings.

Anyone can play any role at anytime.







PDF or Paper Document





Machine Readable Document



Citizens have their own documents as PDF or Paper Document.

This is for human beings to look at it.

All the documents are converted into .idstack format by Extractor and given back to Owner with Digital Signature.

This is for automatically processing by the computers.



Design



Machine Readable Document



All the documents are converted into .idstack format by Extractor and given back to Owner with Digital Signature.

This is for automatically processing by the computers.

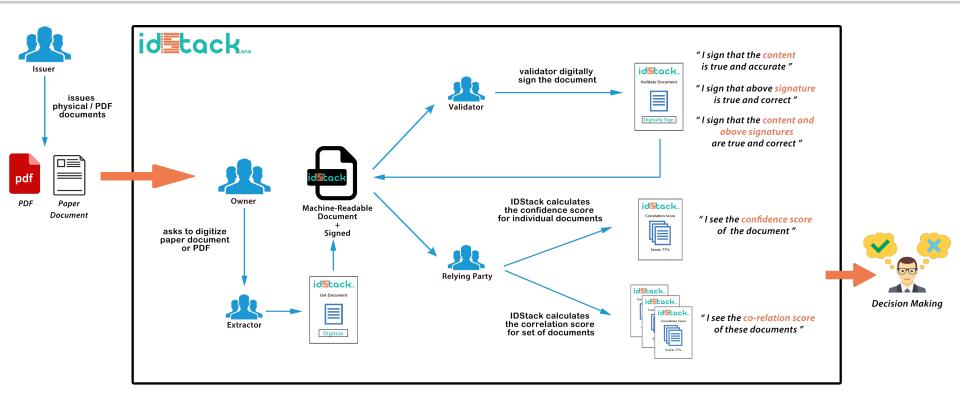
Design

id tack...

	1 - (
1 - {	2 - "meta_data": {
2 - "meta data": {	3 "name": "IDStack",
	4 "version": "1.0", 5 "document id": "a8heEdgg",
3 "name": "IDStack",	6 document_type": "passport",
4 "version": "1.0",	7 * "issuer": {
	8 "organization": "department of immigration and emmigration",
	9 "email": "controller@immigration.gov.lk",
6 "document_type": "",	10 "url": "http://www.immigration.gov.lk/usr/local/idstack/extractor
7 • "issuer": {	/cert/pub/9e18a0b3-45cd-41e9-81a2-073e91e0c1b0.cer"
8 "organization": "",	$11 $ } $12 $ }
	12 // content": {
9 "email": "",	14 "university": "University of Moratuwa",
10 "url": ""	15 "faculty": "Engineering",
11 }	16 "degree": "BSc Eng Hons",
	17 - "name": { 18 "surname": "Dangalla",
12 },	18 surname : Dangalla , 19 "initials": "DADJS"
13 "content": {},	
14 - "extractor": {	21 "index_no": "130094R",
	22 - "address": {
15 "id": "",	23 "linel": "9A, De Mel Watta Road, Koswatta",
16 - "signature": {	24 "line2": "Nawala" 25 },
17 "message digest": "",	25 J, 26 "nic": "937741180V"
	27 }
18 "url": ""	28 - "extractor": {
19 }	29 "id": "b178dad3",
20 },	30 - "signature": {
	31 "message_digest": "MIAGCSqGSIb3D0EHAqCAMIACA0ExDzANBqlqhkqBZQMEAqEFADCABqkqhkiG9w(
21 - "validators": [winderspecture and a standard and a winderschaft and a standard and a
22 - {	32 "url": "http://www.mrt.ac.lk/usr/local/idstack/extractor/cert/pub
23 "id": "",	/9e18a0b3-45cd-41e9-81a2-073e91e0c1b0.cer"
	33 }
24 - "signature": {	34 },
<pre>25 "message_digest": "",</pre>	35 - "validators": [36 - {
26 "url": ""	37 "id": "lued5eg",
	38 - "signature": {
27 },	39 "message_digest":
28 "signed_content": "",	"MIIFRjCCBC6gAwIBAgIRAMxtHVlq5YH597HXLW01nl4wDQYJKoZIhvcNAQELH
29 "signed signatures": []	wgZAXCZAJBGNVBAYTAkdCMRswGQYDVQQIExJHcmVhdGVyIE1hbmNoZ", 40 "url": "http://idstack.one/gtzKCGE0xH87M9F"
	40 "UII:: http://latack.one/gtzkCGEUXH8/M9B" 41 },
30 }	42 "signed content": "true",
31]	43 "signed_signatures": []
32	44 }
	45
· ·	46 }



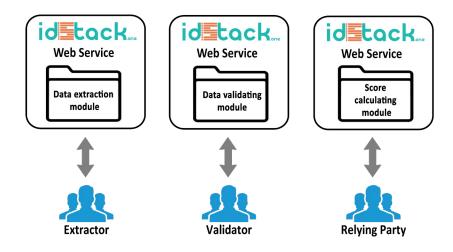
idetack..., The Common Protocol





id tack

Architecture



> Web Service

Spring Boot REST API with 3 modules mapped into 3 user roles

Data extraction module Data validating module Score calculating module

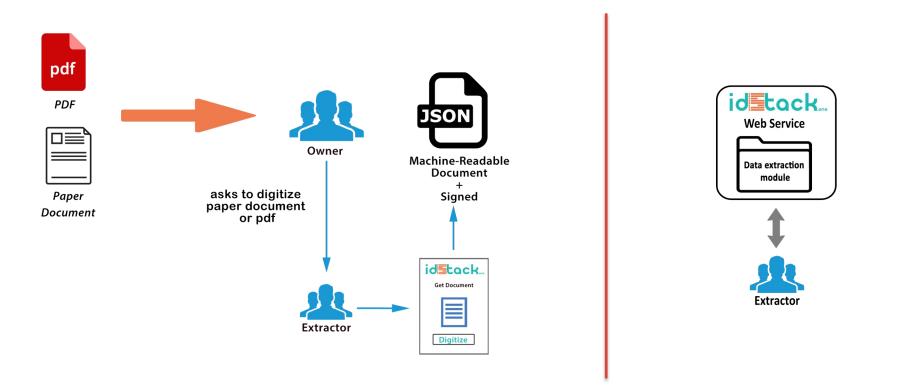
Extractor Validator Relying Party

> Client Applications

Desktop applications built using Electron API for 3 user roles

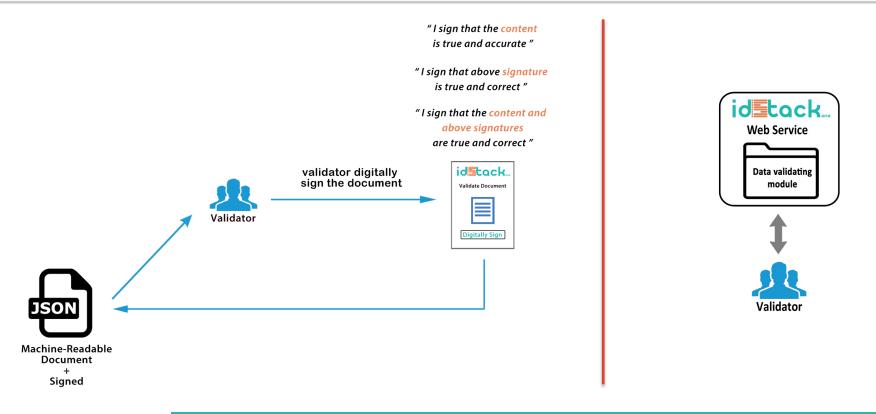


User Role : Extractor Module : Data Extraction Module



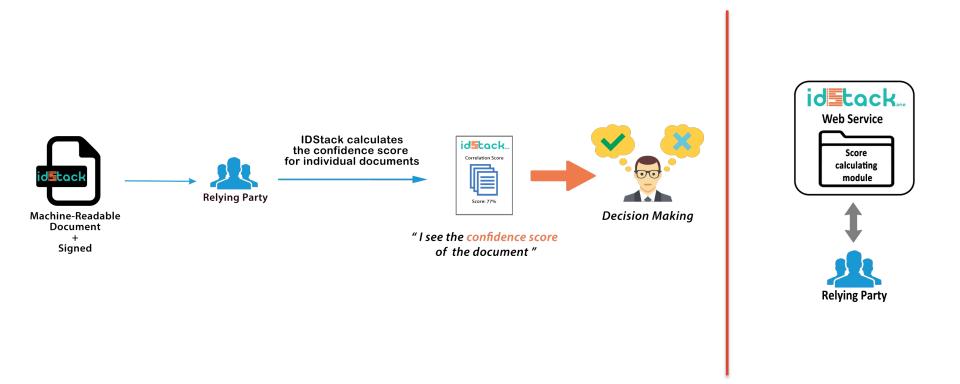


User Role : Validator Module : Data Validating Module



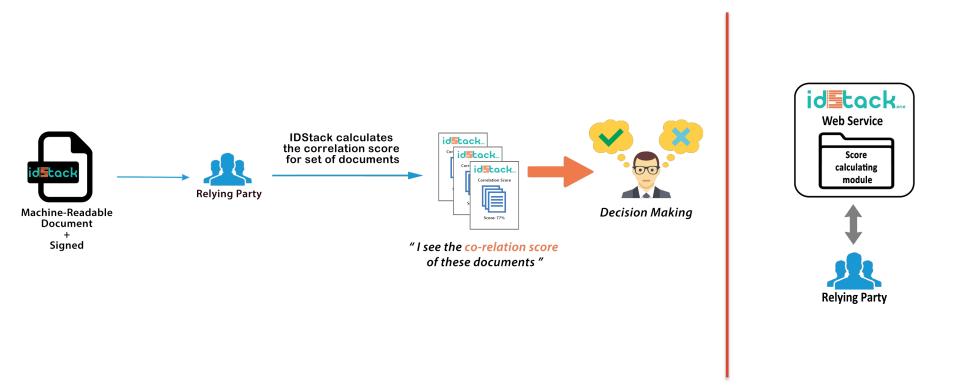


User Role : Relying Party Module : Score Calculating Module





User Role : Relying Party Module : Score Calculating Module







- On average, 3.1 days are added to most processes in order to collect physical signatures
- eSignatures reduce document turnaround time by 80%

Source: AIIM White paper study http://www.aiim.org/pdfdocuments/MIWP-DigitalSignatures-2013.pdf



The Electronic Transactions Act (ETA) No. 19 of 2006 (Section 7) gives e-signatures the same legal weight as traditional hand-written signatures.

Source: Electronic Transaction Act, No.19 of 2006 (Section 7)

In budgets for the years 2016 and 2017, the government allocated LKR 15 billion to implement policy of digitalizing the economy.

Source: Budget Speech 2017, Ministry of Finance





Source: Accepting E-Documents with E-Signatures, VERITE Research (February 2017)

Need of id tack...



Source: The Global Enabling Trade Report 2016, World Economic Forum TABLE 1: TIME TAKEN (IN HOURS) TO COMPLY WITHDOCUMENTATION REQUIREMENTS FOR INTERNATIONALTRADE

	To Export	To Import
Singapore	4	
UAE	6	37
Malaysia	10	10
Oman	31	24
India	61	67
Pakistan	62	153
Sri Lanka	76	58

Source: World Bank, Doing Business Index 2016

Need of id tack...e

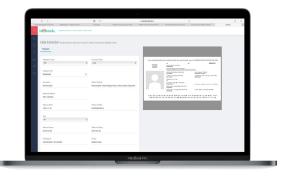
ICTA implement e-Document Attesting System at Ministry of Foreign Affairs



The Information and Communication Technology Agency (ICTA) has taken another step towards creating a digitallyempowered nation by implementing an electronic Document Attesting System (eDAS) at the Ministry of Foreign Affairs (MFA).

Source: http://www.ft.lk/article/596399/ICTA-implements-e-Document-Attesting-System-at-Ministry-of-Foreign-Affairs (February 2017)





Ming-Ampeter Strack Spherican - Mach	dotte: Parker Costs Study	ting active. Adde Partier Consolite Rp., 375-071	Ida Ceates Stud. American Atla Ceates.	_00_
idStock.	NUMBER AND ADDRESS			
0				
Document Verifica	tion way according to apply a part of			
// Select Document	Conce signed person	tan i		
				_
Persont Type:	**	Courtly Code:	004	
Persport No:	NUMBER			
53 Grane	CLM7DN	Other Names	HLARY DUPP	
Network Datase	SRILANON			
tex of time	1878-11-01	Pace-of Britts	KURUNDGALA	
0 M				
Deter of Tessee	201040	Date of Digity	2010-04-05	
Polesson	UNIVERSITY \$10,00NT	- 10 MK	96561+196V	
				_
SIGNATURES				
Clarin Falsanautra				_
Careto Danal Care				28 Pyo, 12 30AM
Extension Explosion April 5 10 (01.01)	mitz Gwill			
96.0				

C X	0	Onable Coult arbitra and exe.	i less prais			Angenuen kinde Cestin.	. 0 0
idStock.		Deaths Coultaring and rec.	Able for the Compiler for	C PERMO	Perfe Ond.	Anazon.com kolde Celafile	COurt
Exhibits Despite schlander Bet	ar	20.33, 81.3094	ngunom. Individual da	on for the Occurrent occurrent score :		ed according to the algorithmic and co	fort that early with
- SCOPE Scowley be insured. The scow	is prevaled according to the signature	a and sortiers that worky with	Same	0	Gr. Hanasingha I Panasingha GR. RANDO Ran	hammingo kasar Dramantas (aganam	
Individual document score :		-77 30	MC	\odot	CF 80001981		
Correlation document score :	energie Annihisp Geor Danobi 194	Inpent	Address	-74		Mar Real, Konnella	
× 0 ···	NYS foreste Sea sev		Genow DOB	-	(* 1999) (* 1990)		
A00000 -	Destroy		temp	-			
Desir (m				-22.69			
00 (m)							
Average							



Extractor App



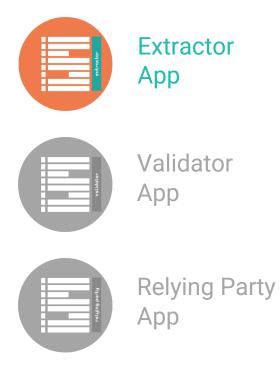
110

Validator App



Relying Party App





		0		ii idstack.github.io		C	A
Biling Management Console	DigitalOcean - Iditack-ubuntz		Creative Cloud pricing and me	Adobe Partner Connection Por	21% Off Adobe Creative Cloud	Amazon.com: Adobe Creative	iDžtack +
Passort Type Passort Type Passort Type Passort No Ne66335 Surrame Ranatinghe National Status Stil LANA Date of Brith 1991-11-01	 (a) a select sau from document or order in (b) (c) (c)<!--</th--><th>Country Code LKA Other Names</th><th>age Kasun Dhanushka Geyaneth</th><th></th><th>P1 Sumple significant / Summer MARCHIER, Discussion / Discussion Researchier / Discussion / Discussion Researchier / Discussion / Discussion Researchier / Discussion Control / Discussion / Discussion / Discussion / Discussion Control / Discussion / Discussio</th><th>aduati (Sigu dia) (Hantson Lawinette NLDENT Unggitted ana / gam, ana gyi, m., unit / O No. 522211599 patrona attac anggi gunat antihana / hatar appan</th><th></th>	Country Code LKA Other Names	age Kasun Dhanushka Geyaneth		P1 Sumple significant / Summer MARCHIER, Discussion / Discussion Researchier / Discussion / Discussion Researchier / Discussion / Discussion Researchier / Discussion Control / Discussion / Discussion / Discussion / Discussion Control / Discussion / Discussio	aduati (Sigu dia) (Hantson Lawinette NLDENT Unggitted ana / gam, ana gyi, m., unit / O No. 522211599 patrona attac anggi gunat antihana / hatar appan	
Sex M Date of Issue 2016-04-25 Profession UNIVERSITY ST	e) TUDENT	Date of Expliny 2016-04-25 ID No 923231196V					

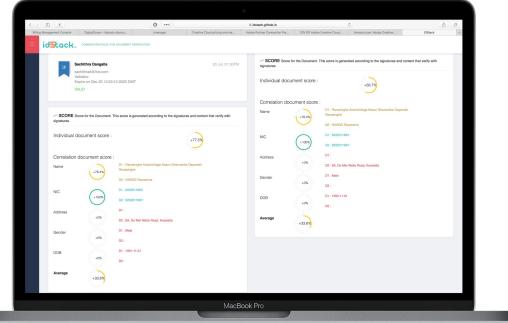


extractor	Extractor App
olidaer	Validator App
Ryad Burnya	Relying Party App

			۲		
C > Diling Management 6		(manager Creative Cloud pricing and me	Adobe Partner Connection Por	C 21% Off Adobe Creative Cloud Amazon.com: Adobe Creative	DStack +
idsto					
Doc	ument Verification	Verify documents using the digital signatures			
0 Sele	ct Document	CHANGE IDStack-signed-passport.json \times			
	assport Type:	PB	Country Code:	LKA	
	assport No:	N6663355			
⊠ _ s	umame:	CLINTON	Other Names:	HILARY DUFF	
	ational Status:	SRI LANKAN			
	ate of Birth:	1978-11-01	Place of Birth:	KURUNEGALA	
S	ex:	м			
	ate of Issue:	2016-04-25	Date of Expiry:	2016-04-25	
_ P	rofession:	UNIVERSITY STUDENT	ID No:	923231196V	
					_
SIGN	ATURES				
	Chamin Wickramarathna chaminb@gmail.com Extractor Expire on Apr 25 10:31:18 2022 GMT VALID	r			26 Feb, 10:30AM
		Maal	Book Pro		
		Maci			



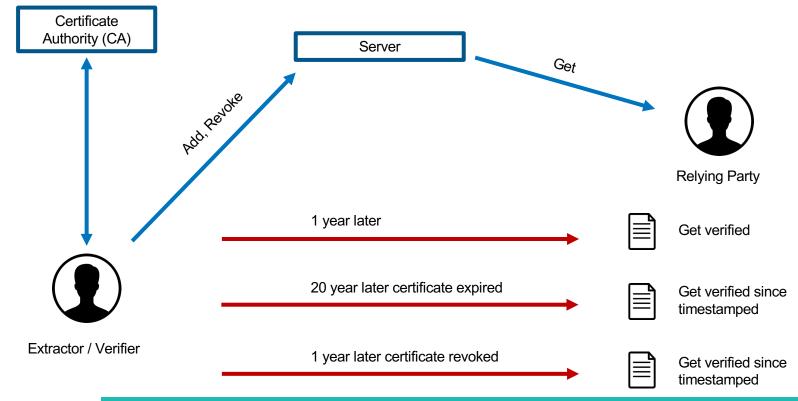
orrector	Extractor App	C C C C C C C C C C C C C C C C C C C
valotee	Validator App	ViLD Vild
Frud Guilty	Relying Party App	Addensi Gander DOB Average





Thank You !







Correlation Score

Super Attributes:

- 1. Name
- 2. Address
- 3. DOB
- 4. Gender
- 5. NIC

Other information (with identical keys) are compared using String similarity measures.



Correlation Score

Person-Identity based attributes can be divided into 4 categories [1]						
	Space-related attributes	Time-related attributes:	Classifying attributes:	Identification codes:		
Example	Address	Date of birth	Gender	Passport No, Social Security No		
Possibility to be updated	High	Very Low	Low	Very Low		
For FYP scope covering Sri Lankan context we selected 4 attributes with weights	Address 0.4	Date of Birth 0.99	Gender 0.99	NIC 0.99		

[1] B. Lisback, Linguistic Identity Matching

id tack....



Correlation Score

Problems:

- 1. Sri Lankan names are long and repetitive: existing name-similarity measuring methodologies can give conflicting results
- 2. Mostly in Sinhala or Tamil: translation to English can have variations
- 3. Different documents represent the name with different attributes and attribute order
- 4. Non-linguistic typing mistakes
- 5. People can change their names

Solution:

idstack

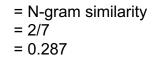
An algorithm that calculates:

- phonetic similarity
- order of name segments
- string similarity

Correlation Score

	Document1	Document2
Example	Ranasinghe Arachchilage K <mark>asu</mark> n Dhanushka Gayanath Ranasin <mark>g</mark> he	K <mark>evi</mark> n Gayanath Ranasinh <mark>a</mark>
Phonetic representation	R525 A622 K420 D522 G530 R525	K870 G530 R525
Identify overlapping name segments	R525 A622 K420 D522 G530 R525	K870 G530 R525
String similarity	Ranasinghe Arachchilage Kasun Dhanushka Gayanath Ranasinghe	Kevin Gayanath Ranasinha

Oder score between doc1 & doc2





Correlation Score

Unique phonetic IDs	Document 1	Document 2	String similarity (Normalized Levenshtein)
R525	Ranasinghe	Ranasinha	80%
A622	Arachchilage	-	
K420	Kasun	-	
D522	Dhanushka	-	
G530	Gayanath	Gayanath	100%
K870	-	Kevin	

String similarity score between doc1 & doc2

= (0.8 + 1) / 6 = 0.3

Total score between doc1 & doc2

= (order score + string score) / 2 = 0 287+0 3/2



Correlation Score

	D1	D2	D3	D4
D1	-	а	b	d
D2	а	-	С	е
D3	b	С	-	f
D4	d	е	f	-

- a = Score between D1 & D2
- b = Score between D1 & D3
- c = Score between D2 & D3 etc.

Correlation score of D1 = (a+b+d)/3



Single document score calculates a weighted average based on the uniqueness weight and determinacy score.

Uniqueness Weights (Wa)			
Super Attributes	NIC	0.99	
	DOB	0.99	
	Gender	0.99	
	Name	0.7	
	Address	0.4	
Minor Attributes	Others	0.1	

Determinacy Scores (Sa)		
Existence of super attribute	0.7	
Existence of signed super attribute	0.9	
Existence of signed minor attribute	0.2	



Appendix 5

API - <u>http://docs.idstack.apiary.io</u>

Sample JSON

Student ID - http://www.jsoneditoronline.org/?id=dd533c47497c94e42e7aabc40e7ab547

Passport - http://www.jsoneditoronline.org/?id=f85dbe5a204a0fa2f9cfb50bfaf10de1



FYP Scope	Input	- Digital PDF & Paper document without OCR
	Language	- English
	Template	- Use predefined set of templates (Student ID, Transcript, Passport, Driving Licence)
Future Work	Input	- Physical Documents with OCR
	Language	- Sinhala & Tamil



A reviewed concept

A concept we use

Similar products [some use cases]

A product we use

idstock

Blockchain Technology - Bitcoin [1], Ethereum [2] Digital Signature DocuSign [3], Stampery [4], ShoCard [5], BlockSign [6]

DocParser [7]

[1] S. Nakamoto, "Bitcoin A Peer-to-Peer Electronic Cash System", 2017. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[2] "Ethereum Project", Ethereum.org, 2017. [Online]. Available: <u>https://www.ethereum.org/</u>.

[3] "Don't settle for less. Choose the #1 eSignature brand.", DocuSign, 2017. [Online]. Available: https://www.docusign.com

[4] L. I. C. G. Adán Sánchez de Pedro Crespo, "Stampery Blockchain Timestamping Architecture," [Online]. Available:

https://s3.amazonaws.com/stampery-cdn/docs/Stampery-BTA-v5-whitepaper.pdf.

- [5] "ShoCard | Identity for a Mobile World", Shocard.com, 2017. [Online]. Available: https://shocard.com/
- [6] "BlockSign", Blocksign.com, 2017. [Online]. Available: https://blocksign.com/
- [7] "Docparser Extract Data From PDF to Excel, JSON and Webhooks", Docparser.com, 2017. [Online]. Available: https://docparser.com