New Times Digitization Project Proposal : Phase 1

Presented by Software-Wolves from E-SW

Software-Wolves

By E-SW

Project Summary:

This project aims to completely digitize all of The New Times last 25 years worth of newspapers; news clips, articles, internal correspondence, past news reports, sales figures, staff info/salaries, taxes, bills, ads, prior reader letters, etc. for The New Times.

We aim to complete the digitization by January 1st of 2021. The digitization must be complete, accurate, secure.The data must be searchable by: date, author, subject, headline, geography, and string literals within text.If the projected completion date is not met, the client has agreed to allow for an extension to January 1st 2022 pending a potential renegotiation of terms.

For any inquiries, please contact any of the project managers:

Dechen Chuteng	916-131-1258	dechchu@e-ws.net
Scott Stage	916-123-9876	sstage@e-ws.net

Feasibility:

Stage 1: This stage handles the actual organization and digitization of existing documents from The New Times. This will involve physically scanning any and all documents that NT requests to be digitized. This will include news clips, articles, internal correspondence, past news reports, sales figures, staff info/salaries, taxes, bills, ads, prior reader letters, etc. The step will require that our team purchase a high-volume, large format scanner. We will have two of our team members in charge of the scanning process. Our team will take extra care to make sure all physical documents are well cared for and preserved during the scanning process. By our estimates, the scanning process will take no more than 240 hours (30 work days)

Stage 2: This stage can run partially in parallel with the scanning stage, is to develop the software needed to process the scanned documents. All scanned documents will be saved in word format. The team in charge of scanning will save all documents in a database according to the category being scanned at any one time. For example, all tax documents will initially be stored in the taxes portion of the database, etc.

Stage 3: All documents that now reside in a database will then be integrated into a self contained program that will allow NT to search for a desired document according to keyword, category, date, etc. Sensitive documents will be sectioned off and are only allowed to be accessed by qualified individuals within the NT organization. All files will also be backed up in an offsite location, ensuring that all files will be secure in case of any negative unforeseen circumstances.

Our company specializes in the development of custom software solutions that offer high reliability, accuracy, security, usability, and portability. We have no doubt that the digitization of The New Times' documents is something that the Software-Wolves can complete within the required deadline, and will do so under budget.

The NT Contact Information:

John Jonah Jameson Jr. *Editor-In-Chief at The New Times*	JJJJ@jmail.com	916-123-4567 ext. 892

Product Life Cycle:

We will be using the four phase software Product Life Cycle(PLC) model to manage and track the progress for the development of this digitization project. We will first begin with the Exploration Phase: perform research to identify solutions to digitize the data(News clips, Articles, Internal correspondence, past news reports, sales figures, staff info/salaries, taxes, bills, ads, prior reader letters, etc) for The New Times. Next is the Planning Phase; where we will define the project scope through requirement gathering and analysis. Our team will create an outline for deadlines to complete all promised features, get the needed fundings for this development, and make technology integration a smooth transition for the NT. Next is the Development Phase, where we will implement the software(detailed coding) for digitizing the company's old and future data, create documentation, and create pre-production sample software releases for early beta users. Finally, the Refresh Phase will take place. Our team will perform further revisions and fix any existing bugs in the software after the Development phase. We will be reviewing with NT in regards to any other key features they would like us to add or remove. Reaching the end of this phase, Software-Wolves will ensure to include any software or hardware updates needed to keep this software running on its own.

Review Dates for the following phases:
- Exploration Phase: 1/17/2020, 1/31/2020
- Planning Phase: 2/14/2020, 2/28/2020
- Development Phase: 03/02/2020, 08/14/2020
- Refresh Phase: 09/17/2020, 11/16/2020

Material Review:

The categorization of the required document will be an all hands on deck affair. The sooner we can get them organized, the sooner we can start the scanning process. Assuming NT had a reasonable system of organizing their documents, this process should not take long. With our entire team working towards getting everything organized, we should have the documents ready to scan within the first 30 days.

Scanning Procedure:

We will put two members of our team in charge of the scanning process. The documents will be scanned into a database according to the document type being scanned at the time. Average newspapers print between 40-75 pages per day. That number over the span of 25 years would account for roughly 626,000 pages. Based on commercially available high volume, large format scanners, which can scan roughly 120 pages per minute, it should take roughly 87 hours to scan the estimated 626,000 pages worth of newsprint. This is a best case scenario however, so to be safe, we shall double the allotted time to make sure that we account for extra time needed.

Assuming sales reports are per day, this would account for roughly 9300 sales reports which would take about 3 hours to scan. Again, we'll double this estimate.

Similar rationale for internal correspondence, previous staff information and salaries, taxes, bills, ads, former reader letters.

The majority of the time needed to scan documents would be spent scanning news clips and reports. Even with an over exaggerated estimate for the time needed to scan everything, it should be done in under 45 working days. This leaves ~320 days left for the development of the software needed to digitize all of the data.

We will update NT periodically during the scanning process as well as let them know when the process is completed.

The one year deadline will not be an issue.

Schedule:

Jan 1 - Mar 4: Exploration and Planning
March 4 - April 4: Organize and categorize all documents in preparation for scanning
April 5 - June 19: Scan all documents into our database according to document type
April 24 and June 19: Inform NT of the scanning progress and completion
April 5 - August 13: Development of required software solution
August 14: Present NT with a working prototype and gather feedback and recommendations
August 17 - September 18: Implement any needed changes
August 14 - December 20: Test and verify functionality

Month	1	2	3	4	5	6	7	8	9	10	11
	Mar '20	Apr '20	May '20	Jun '20	Jul '20	Aug '20	Sep '20	Oct '20	Nov '20	Dec '20	Jan '21
Phase One	Organize Documents										
		Scan Documents									
		Development						Test & Verify Functionality			Delivery
						Prototype Presentation	Implement Changes				
		Scan Update		Scan Update							

Environment:

We will be using a large scale scanner machine to scan all old and new printed documents and convert them into files which are easily searchable and located in a MySQL database. We intend to use an integrated development environment to implement the software solution. In our software, we will be storing all the digitized data in our secure database. To further keep the data secure, we will use a third party cloud based storage.

Database

All documents will be digitally backed up to an offsite location immediately upon scanning them into our own database. This will ensure that all documents will be safe in case of a system malfunction or catastrophic event. In the event that any data is corrupted, those files in question can be reuploaded to the NT system. Additionally, all hard copies will be stored categorically in a fireproof storage container on our premises.