# **Ata de Reunião da NE-Soft-Factory**

Data 31/07/2018

Veículo: Skype

Início: 21:00 h Término: 21:59 h

Duração: 00:59 h

Participantes: Alfeu, Maria do Carmo, Helton, Fernando, Erick, Antônio Amorim, Antônio Filho

**Assuntos Tratados**:

* Alfeu começou falando sobre a instalação do Jira, a qual está em ordem agora;
* Alfeu falou sobre as heurísticas que Fernando, Amorim e Erick iriam terminar no domingo, mas, o pessoal falou que ficou trabalhando na instalação do ambiente, resolvendo alguns problemas de configuração e integração de ferramentas;
* Alfeu falou que esteve pesquisando e estudando o OCR, e que a estrutura, forma como converte o texto e grava no documento, é desestruturada, ele gera o pdf com duas colunas e os softwares de OCR não entendem. As tabelas são capturadas de forma desorganizada e são eliminadas por alguns dos softwares de OCR. Existem outros pontos mais graves, como o diário der Parnaíba que tem uma perda grande de caracteres (5%), que quando o OCR não entende o caractere, o substitui por outro, mudando o significado do texto. O pessoal tem que pesquisar, esses problemas podem inviabilizar o projeto no formato em que foi pensado. Se deveria criar uma variação para esses casos especiais, mas o tempo é curto para fazer isto;
* Erick falou que o D.O. de Teresina é todo texto, é só ler o arquivo;
* Alfeu falou que mesmo assim tem tabelas que dificultam a leitura pelo OCR;
* Alfeu disse que o importante é Antônio Filho pensar em outras demandas do TCE-PI para mudar o projeto original;
* Alfeu falou que a publicação com tabelas quebra todo o trabalho do OCR;
* Amorim falou se não poderia transformar em html no lugar de texto;
* Helton perguntou se poderia desconsiderar as tabelas, mas Antônio Filho falou que as nomeações sempre vêm em tabelas, e falou que se fizer só a fonte de Teresina, vai tirar o valor do produto;
* Helton falou que todo mundo deveria fazer mais testes com OCR, com base no que Alfeu fez;
* Antônio Filho falou para fazer o teste no PDF que Alfeu trabalhou;
* Helton falou que se deveria fazer o teste pegando a coluna direita e depois a coluna esquerda do texto para formatar separadamente;
* Helton falou que devemos tomar a decisão de como vai fazer o projeto de agora em diante, até sexta-feira próxima ou segunda, porque se não funcionar, ‘nada feito’;
* Helton falou que Maria e Amorim devem continuar com os documentos das aulas (Heurísticas, dimensões, plano de projeto e status repórter) e deixar os outros com os testes de OCR;
* Todos aprovaram;
* Alfeu falou que o OCR lê o conteúdo das tabelas e converte os dados fora de ordem, deu print no PDF, mandou o OCR converter a coluna, mas o resultado ainda não é bom;
* Alfeu disse que o Adobe consegue identificar as colunas do meio, as figuras e etc., mas, ele não ignora e aí se perde, a qualidade é ruim;
* Alfeu falou que as tecnologias OCR hoje para texto ainda não alcançaram um nível bom. Ele está muito desanimado;
* Helton falou para Alfeu fazer uma demonstração do resultado da pesquisa com o texto e o resultado de cada ferramenta, para comparação;
* Helton falou que tendo uma fonte mais fácil, padronizada, que não tivesse tabela seria ótimo, mas a maioria das fontes têm tabelas, tentou fazer teste para nomeações, mas, não teve tempo de pesquisar. Ele falou que é possível transferir as tabelas para excel, mas ainda não tentou, não encontrou também ferramentas para converter textos complexos como os dos D.O.;
* Fernando falou que fez demonstrações com ferramentas, mas foi para textos simples;
* Alfeu falou que além de trabalhoso é muito complexo, que vai solicitar ao professor Felipe a informação sobre algum grupo que já tenha trabalhado com OCR;
* Alfeu pede para Antônio Filho ver se tem um plano B para o projeto;
* Amorim perguntou se existe e-mail da fábrica e Alfeu respondeu que tem, mas que não acessou;
* Amorim falou sobre as tarefas dos entregáveis para confirmar: similares, heurísticas, atualização do processo de testes;
* Amorim falou que falta definir o processo de testes da fábrica no template de Ana Cláudia, e que para o status repórter a professora iria mandar o template por e-mail;
* Fernando solicitou que todos olhassem o que ele mandou por slack;
* Alfeu pediu para todos reservarem tempo para pesquisar o OCR, senão vai inviabilizar o planejamento inicial;
* Amorim pediu para Alfeu disponibilizar o PDF mais complicado para todos estudarem a partir dele;
* Alfeu falou que poderia procurar uma fonte mais padronizada, mas Antônio Filho falou que isto não agregaria valor ao produto, trabalhando com uma fonte apenas;
* Alfeu falou que Antônio poderia levantar uma outra necessidade de software do TCE-PI para substituir o projeto. Alfeu lembrou que não dá para desenvolver um OCR em pouco tempo;
* Amorim sugeriu que ele e Maria focassem nos entregáveis pelo pouco tempo que têm e que os outros focassem na pesquisa de OCR;
* Antônio solicitou o PDF para mostrar aos auditores as dificuldades e as alternativas;
* Amorim estava pesquisando um drive, mas não teve muito sucesso;
* Alfeu falou que vai colocar no Jira, o backlog do produto e se tudo der certo irá planejar a sprint. Falou também que o Jira substitui o Trello;
* Alfeu irá fazer o calendário das reuniões e enviar.