

ISKO Singapore – Intelligent Document Processing with Maish Nichani
April 14th 2022

Edited notes of chat channel

Maish: There is a power law for enterprise IT delivering digital services - from high transaction volume applications to mixed manual, partially digital, shadow IT, for low volume transactions. Traditionally, low volume transactions have not been cost effective to address, hence the shadow IT approaches. Only recently (because of pandemic + digital transformation) does it start to become cost effective to start addressing the long tail. Now we are seeing a vast expansion in IT services / app development in low code intelligent automation tools. Low-code automation tools are pre-fabricated building blocks for application building, which makes low volume applications more affordable. This is provoking big shifts in IT capability requirements within enterprises.

Maish: With structured data - elements have fixed locations and are processable according to their location in the database tables. Content in documents is unfixed and not so easily processed. IDP is about taking unstructured content in web pages, documents, images, etc, and converting it into structured forms so that it can be processed and acted upon.

Maish: Digital workers are bots that can use intelligent automation tools to perform manual, repeated work previously assigned to humans. Key areas include computer vision, sentiment analysis, learning and improving from corrections, smart workflows.

Bill: My experience with Customer Service bots is not very positive. They fail quickly.

Kah King: Yes, I agreed some bots are very good and powerful but not most deployed in organisations here. Also, where the problem deals with fast changing situation like COVID, they don't help much. They work well with those environment where policy and knowledge don't change much over time.

Patrick: Question about potential governance issues in opaque, low-code / no-code systems - e.g. security, privacy, structural/social bias, etc.

Bill: When systems work with documents, people become very concerned because the documents are created in an uncontrolled environment. The intent (e.g. informality) of a given document may be inconsistent with the way the extracted data is used.

Maish: IDP allows organisations to process content from less “structured” forms of templated documents. Document classification identifies what kind of document it is, and then it can infer key types of information from that document type, using text analytics that has been pre-trained on very large document sets. Some incremental training is still needed on a few hundreds of local documents to tune the tool to local content.

Maish: Positional analysis vs. content analysis - the text analytics tools infers not by position but by understanding what the elements are, e.g. roles, personal names, organization names, telephone numbers, addresses, etc. Training sets are very important on a given document type. Nowadays, the AI/ML space has advanced using a common operating model, so new types of document are incremental on that base, so we may only need a couple of hundred documents in a training set now for a new type of document. Human review stages can be interspersed based on the tool's confidence threshold / degree of confidence in the assessment.

Kah King: In my view, if it is used to analyse one annual report, I would rather read the annual report myself as I will know what to look for. However, if I can take 4 competitors' annual reports and my company into the IDP, and it can have an analysis report based on various area or fields (in terms of who does better in each area, what are the market trends, how is my company positioned in the market, etc.) it will be great. Where is the line between IDP and business intelligence?

Maish: The point of an IDP platform is to convert unstructured to structured so that it can be useful for downstream applications, such as business intelligence tools, data processing platforms, workflow engines, analytics tools and so. It is very important to understand what is your objective and what downstream applications will be consuming the structured data, first, before deploying IDP. We work backwards from the purpose and the downstream applications.

Bill: Collecting lots of data before knowing how it will aid compliance is dangerous.

Maish: Training of the platform is progressive. First you teach, then you correct their predictions.

Patrick: Is there potential bias in the background training sets? Is there transparency in how those background training sets have been built and how assessments have been reached?

Maish: Transparency is important to be able to explain the decisions that the AI is making.

Eileen: What advice would you give to orgs that want to go into this?

Maish: don't go for a point solution for a single purpose - go for something that can scale and apply horizontally - choose a platform over a point solution – something that offers the ability to solve a problem across business lines – e.g. onerous paper trails on key activities. Point solutions may work once but don't scale, and therefore the ROI of IDP is reduced.

Maish: Think about your downstream applications. Pick a strong use case, run an MVP (minimal viable product) project, to get insights especially into digital and business transformation spillover effects, and that can scale. For scaling: the key concept is Centre of Excellence - create a team, leadership, commitment, resources to orchestrate the use of IDP across business functions. And don't forget governance.