

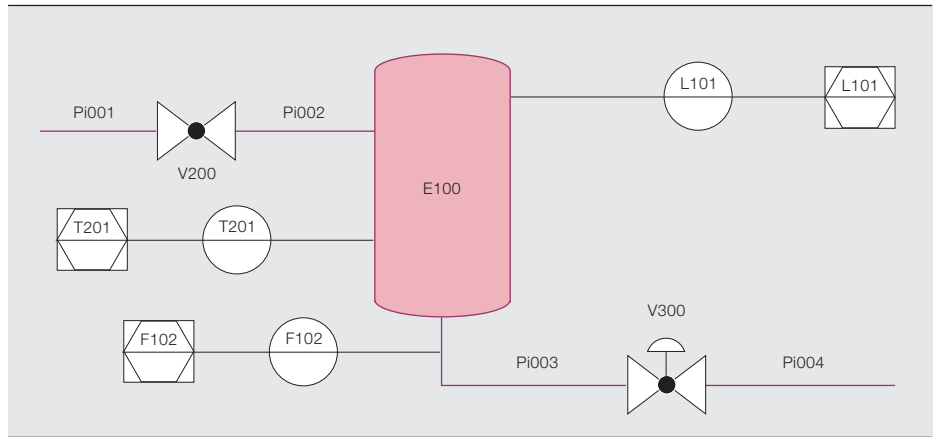


Vers le papier connecté

Un projet de recherche pour extraire des descriptions orientées objets des schémas de tuyauterie et d'instrumentation

ESTEBAN ARROYO, ALEXANDER FAY, MARIO HOERNICKE, PABLO RODRÍGUEZ, STEVE ROYSTON – Dans l'industrie des procédés, la documentation peut prendre des proportions impressionnantes et emprunter de multiples formes, du papier traditionnel aux représentations numériques élémentaires. Regrouper cette information éclectique dans un même format accessible et à jour, notamment en vue de moderniser un site, est une tâche titanesque.

ABB a donc cherché un moyen efficace d'extraire automatiquement et de façon cohérente des modèles de ces données techniques. Un projet de recherche mené avec l'université Helmut Schmidt de Hambourg (Allemagne) a débouché sur une méthode couplant reconnaissance optique et analyse sémantique pour convertir des schémas de tuyauterie et d'instrumentation (TI) en modèles orientés objets.



Pour y remédier, ABB et l'université Helmut Schmidt (HSU) ont mis au point des méthodes d'extraction de représentations orientées objets (OO) à partir des documents de conception existants, en particulier les schémas TI [2].

Des précédents travaux de recherche ont montré que l'emploi de schémas TI pouvait accompagner les différentes

tâches qui jalonnent le cycle de vie de l'usine, telles que les études de danger automatisées HAZOP (*HAZard and OPerability*) [3], la détection de modèles de conception [4], l'obtention de modèles de simulation [5], la recherche de défauts et le diagnostic [6]. Les méthodes existantes ont toutes un prérequis: les schémas TI doivent être décrits au format OO, normalisé CEI 62424 CAEX/AML [7,8] → 4 ou ISO 15926 [9], par exemple. Or ce n'est malheureusement pas le cas dans le *process* où la majorité des schémas TI est archivée sous des formes inintelligibles pour l'informatique.

Des précédents travaux de recherche ont montré que l'emploi de schémas TI pouvait accompagner les différentes tâches qui jalonnent le cycle de vie de l'usine, telles que les études de danger automatisées HAZOP (*HAZard and OPerability*) [3], la détection de modèles de conception [4], l'obtention de modèles de simulation [5], la recherche de défauts et le diagnostic [6]. Les méthodes existantes ont toutes un prérequis: les schémas TI doivent être décrits au format OO, normalisé CEI 62424 CAEX/AML [7,8] → 4 ou ISO 15926 [9], par exemple. Or ce n'est malheureusement pas le cas dans le *process* où la majorité des schémas TI est archivée sous des formes inintelligibles pour l'informatique.

Donner du sens à une mosaïque de points

La méthode développée par ABB et HSU extrait des modèles OO des documents textuels et graphiques que constituent les schémas TI → 1, par la reconnaissance des textes, des formes symboliques et des liaisons entre éléments. Elle part du principe que les documents d'entrée ont quatre composantes fonda-

mentales: des formes paramétriques, des formes non paramétriques, des annotations textuelles et un contenu sémantique sous-jacent → 2.

Le processus de conversion suppose que les schémas TI soient stockés dans des fichiers PDF/bitmap ou PDF/SVG dont on peut extraire les images matri-

Les documents d'entrée ont quatre grandes composantes: les formes paramétriques, les formes non paramétriques, les annotations de textes et le contenu sémantique sous-jacent.

cielles. Pourtant, dans certains cas, il faut passer par des étapes préalables pour obtenir une image matricielle, surtout quand il s'agit d'archives papier ou de fichiers CAO → 3. Ensuite, la méthode exécute trois procédures pour générer le modèle interprétable par l'ordinateur:

- Reconnaissance optique OSR pour les symboles, OCR pour le texte;
- Analyse sémantique: interprétation de la connotation fonctionnelle des formes graphiques pour améliorer la procédure de reconnaissance et les capacités expressives de la description de modèle;
- Représentation et vérification: description des informations saisies sous forme de modèle OO, puis contrôle visuel.

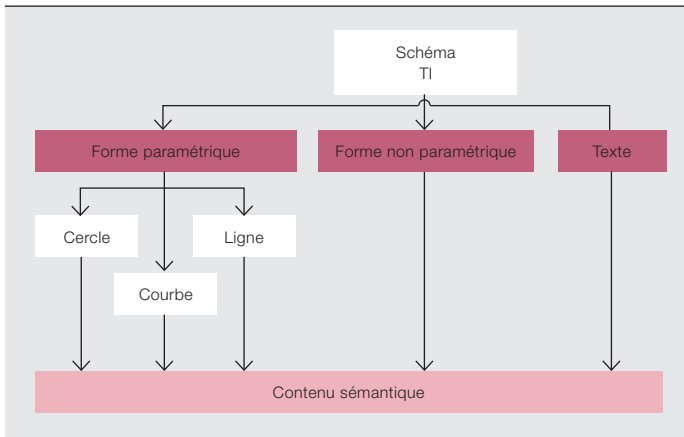
Le gigantisme et la complexité des usines de transformation, de même que la multitude de modifications et d'ajouts effectués tout au long de leur cycle de vie, font exploser les volumes d'informations engrangées: modèles structurels, descriptions fonctionnelles, schémas TI, spécifications d'équipements. Même si cette documentation est un précieux atout pour la planification et l'exploitation d'un site, il est toujours difficile de la localiser au bon moment pour en extraire la bonne information [1].

Numérisation par scanner, stockage et indexation à la main sont aujourd'hui des pratiques courantes dans l'industrie pour convertir les classeurs papier existants en une bibliothèque technique informatisée. Les fichiers obtenus renferment des informations non structurées dont le contenu sous-jacent, faute de répondre à un modèle de données, se prête difficilement aux tâches d'automatisation.

Photo p. 65

La documentation technique dans l'industrie des procédés revêt de multiples formes hétéroclites (papier, fichiers électroniques, etc.). Comment convertir le tout dans un format standardisé orienté objet ?

2 Composantes d'un schéma TI type



Reconnaissance optique

Ce processus déploie des techniques éprouvées de traitement de l'image pour identifier les formes géométriques et les identifiants textuels. On utilise pour cela deux méthodes récursives : la reconnaissance de symboles OSR et la reconnaissance de caractères OCR.

Symboles (OSR)

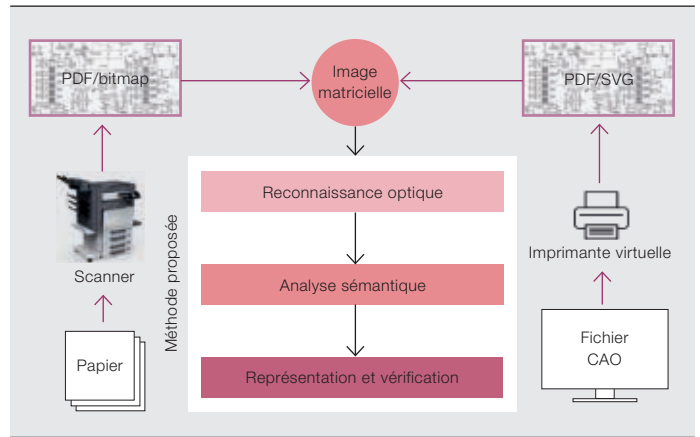
Première étape, les formes « non paramétriques », c'est-à-dire constituées de multiples courbes comme la cuve notée *E100* et la vanne *V300* → 1, sont localisées et comparées à des bibliothèques prédéfinies de symboles. Une bibliothèque peut répertorier, par exemple, tous les équipements de production ou les jeux de symboles usuels représentant des appareils ou processus spécifiques

exhaustive des correspondances possibles : la rotation, la mise à l'échelle et l'occlusion.

Des scores de ressemblance entre candidats identifiés et objets prédéfinis (modèles) permettent de définir plusieurs niveaux de confiance dans la méthode et d'en évaluer la fiabilité. Les objets reconnus sont supprimés de l'image pour faciliter la poursuite du processus.

Vient ensuite l'identification des symboles « paramétriques », c'est-à-dire des formes mathématiques de type cercle ou ligne, telles que la conduite *Pi002* et le capteur *L101* → 1. Les méthodes de reconnaissance utilisées à ce stade sont capables d'identifier les candidats potentiels indépendamment de leur taille et

3 Processus de numérisation



de leur orientation. Il est possible de fixer des seuils de dimensions minimales pour éviter de confondre caractères et symboles. Là encore, la fiabilité de la reconnaissance est réglable selon les scores de correspondance calculés ; de même, les coordonnées de position des objets reconnus sont supprimées pour faciliter l'identification des autres éléments.

Caractères (OCR)

Partant du principe que l'identifiant textuel d'un objet figure autour ou dans l'objet, les coordonnées des formes identifiées servent à générer des fenêtres de recherche → 6, de taille habituellement proportionnelle aux dimensions de l'image

4 Terminologie

AML	<i>Automation Markup Language</i> : format de données neutre intégrant les principales informations utilisées depuis la conception et la planification jusqu'au lancement de la fabrication.
CAEX	<i>Computer-Aided Engineering eXchange</i> : format d'échange de données techniques utilisé dans l'ingénierie des systèmes d'automatisation industrielle.
CAO	Conception assistée par ordinateur
OCR	<i>Optical Character Recognition</i> : reconnaissance optique de caractères
OO	Orienté objet
OSR	<i>Optical Symbol Recognition</i> : reconnaissance optique de symboles
PDF	<i>Portable Document Format</i> : format de fichier qui préserve les polices de caractères, les images, les objets graphiques et la mise en forme d'un document source, quelle que soit la plate-forme matérielle et logicielle utilisée pour l'imprimer ou le visualiser.
SVG	<i>Scalable Vector Graphics</i> : format d'images vectorielles basé sur le langage de balisage XML.
TI	Tuyauterie et instrumentation

Une règle sémantique peut stipuler, par exemple, qu'une conduite d'écoulement est raccordée à un maximum de deux éléments, un à chaque extrémité.

dans les schémas TI. La définition de bibliothèques est non seulement gage de modularité, ce qui optimise les procédures de recherche, mais aussi d'efficacité pour la saisie et la réutilisation des connaissances structurelles au sein de projets. Le critère de correspondance employé se fonde sur les caractéristiques structurelles des formes analysées. Trois propriétés géométriques fondamentales → 5 concourent à la recherche

Au-delà de la détection de connectivité, d'autres règles sont testées pour corriger automatiquement les erreurs et en avertir l'utilisateur.

analysée mais aussi paramétrable par l'utilisateur. Un algorithme OCR y recherche le texte : s'il trouve plus d'un mot-clé ou « étiquette » (*tag*), des règles de nomenclature (expressions régulières, par exemple) sont utilisées pour écarter les chaînes erronées et attribuer une étiquette unique à chaque objet. Les ambiguïtés (telles deux caractères accolés) sont signalées à l'utilisateur pour être levées manuellement.

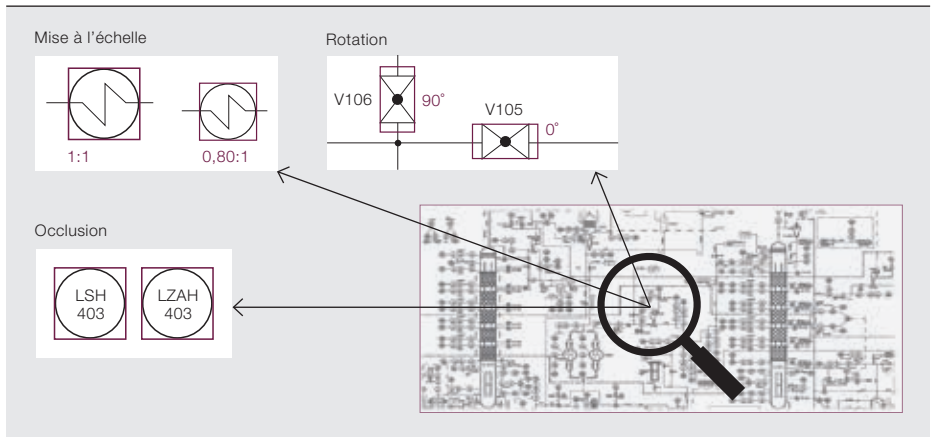
Analyse sémantique

Dans le cadre de notre méthode, on entend par là l'incorporation d'un savoir « métier » spécifique au domaine, en particulier un contenu fonctionnel et structurel, à la reconnaissance des formes graphiques et à leurs interrelations. Entre autres objectifs, cette étape sert à détecter la connectivité et à améliorer les capacités expressives de la modélisation par la définition de types spécifiques d'interface.

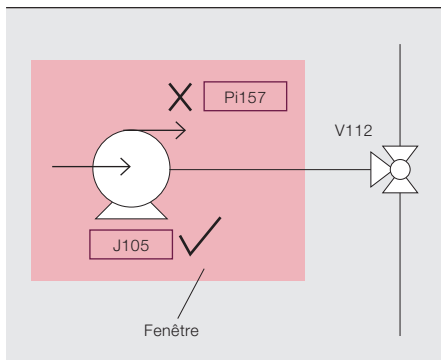
Détection de connectivité

En dessin industriel, les liaisons caractéristiques sont basées sur des segments chaînés multiples ou « polyignes ». La détection de connectivité doit donc commencer par unifier ces lignes contiguës en des éléments à liaison simple. La proximité de ces éléments avec les formes structurelles est analysée pour déduire la connectivité sous-jacente du schéma. Des règles sémantiques sont appliquées pour garantir la cohérence des liens : il est possible de stipuler, par exemple, qu'une conduite peut être raccordée à deux autres éléments au maximum, un à chaque extrémité. Si des instances potentielles ne remplissent pas cette condition, d'autres règles sont testées pour corriger automatiquement les erreurs de détection et en avertir l'utilisateur.

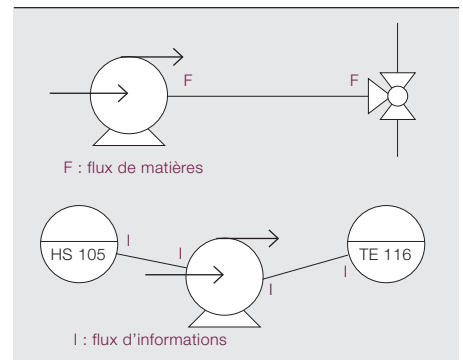
5 Propriétés intervenant dans le processus de correspondance géométrique



6 Fenêtre de recherche et attribution d'identifiants (étiquettes)



7 Types d'interface



8 Exemple de matrice de connectivité

	Cuve E100	Conduite Pi002	Capteur L101
Cuve E100		F	I
Conduite Pi002	F		
Capteur L101	I		

9 Exemple de table de coordonnées

Équipement	Coordonnées TI (X,Y)	Dimensions (L,I)
Cuve E100	(81,123)	(123, 57)
Conduite Pi002	(59,142)	(22,1)
Capteur L101	(105, 126)	(18,18)

Définition de types d'interface

La génération de modèles OO détaillés oblige à distinguer plusieurs types de connectivité. Un exemple : la sémantique d'une liaison cuve-capteur est foncièrement différente de celle d'une liaison cuve-conduite → 1 ; en effet, la première transmet des informations sur des fils électriques, quand la seconde véhicule des flux matières. Cette distinction est cruciale dans plusieurs cas de figure, comme la recherche de défauts et le diagnostic, où l'analyse des causes de défaillance est étroitement liée au milieu dans lequel se propagent les variables perturbatrices du procédé → 7.

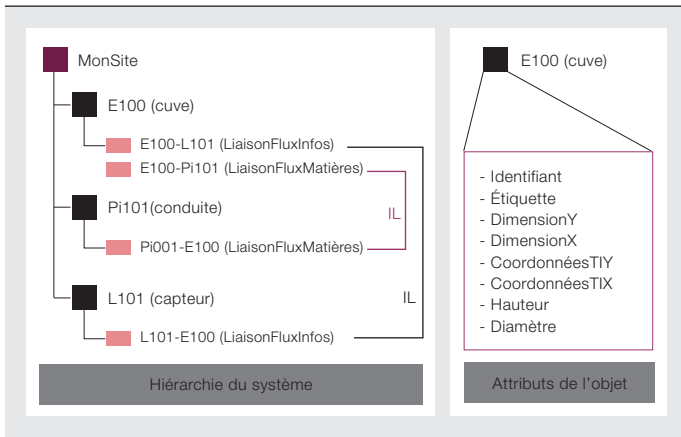
Représentation et vérification

Les informations obtenues par les algorithmes de traitement d'image passent par une étape intermédiaire de description structurelle, avant de donner un modèle OO. Le processus s'achève par un contrôle sur interface utilisateur pour vérifier *de visu* la cohérence des éléments numérisés.

Description structurelle intermédiaire

Pour fournir une première représentation de l'information rassemblée, les liaisons ainsi que la position et les dimensions des éléments trouvés sont stockées dans une « matrice de connectivité » et une table de coordonnées.

10 Exemple de modèle OO obtenu



En première colonne et en première ligne de la matrice → 8 figure la liste des éléments reconnus, répertoriés et nommés (cuve E100, etc.). Les données saisies à l'intersection ligne-colonne représentent les liaisons entre éléments, repérées par une lettre : *F* pour flux matières, *I* pour flux d'informations.

La table de coordonnées → 9 recense la position et les dimensions des objets.

Cette description structurelle intermédiaire peut être codée dans un tableur bureautique (Microsoft Excel ou formulaire C# Windows Forms), consultable à volonté pour obtenir les modèles OO visés. Elle peut également servir à des tâches élémentaires comme la création de nomenclatures ou la réponse à des requêtes sur la présence d'un article ou d'un type d'article donné.

Modélisation objet

De ces données collectées est déduit un modèle OO, notamment CAEX/AML normalisé CEI 62424 [7,8]. On commence par créer un objet pour chaque entrée figurant dans la première colonne de la matrice de connectivité, auquel on ajoute des interfaces : dans l'exemple → 10, deux interfaces sont créées pour la cuve E100, l'une de type F (flux matières), l'autre de type I (infos). La procédure s'achève par l'affectation d'attributs (position et dimensions, par exemple). On obtient ainsi un modèle OO décrivant la connectivité et la sémantique du schéma TI.

Contrôle visuel

Dernière étape, une représentation graphique couleur des informations déduites → 11 est superposée à l'original TI

en noir et blanc. L'utilisateur peut alors visualiser les objets identifiés, les comparer au schéma d'origine et, au besoin, apporter des modifications. Les corrections effectuées dans l'interface graphique sont directement reproduites dans le modèle OO, évitant le recours à d'autres outils de développement. Des avertissements générés au préalable peuvent servir à fixer les priorités de ce dernier pointage.

Le meilleur de la documentation

C'est en améliorant l'exploitation de la documentation existante et en automatisant sa numérisation et sa modélisation que l'on accroît l'efficacité de l'ingénierie des projets de construction et de modernisation de sites industriels. Nombreux sont les secteurs qui peuvent en tirer bénéfice, comme la chimie, la pharmacie, le pétrole et le gaz, l'eau et l'énergie.

Esteban Arroyo

Alexander Fay
Helmut Schmidt University
Hambourg (Allemagne)
esteban.arroyo@hsu-hh.de
alexander.fay@hsu-hh.de

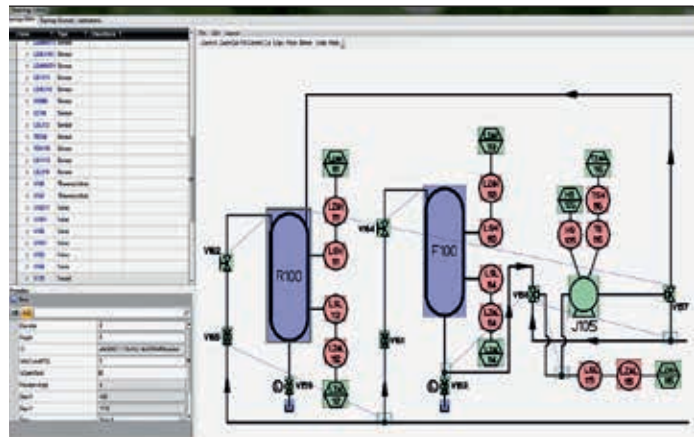
Mario Hoernicke

Pablo Rodríguez
ABB Corporate Research
Ladenbourg (Allemagne)
mario.hoernicke@de.abb.com
pablo.rodriguez@de.abb.com

Steve Royston

ABB Process Automation,
Oil, Gas & Chemicals
St Neots (Royaume-Uni)
steve.royston@gb.abb.com

11 Contrôle visuel



Bibliographie

- [1] Arroyo, E., et al., « Integrating Plant and Process Information as a Basis for Automated Plant Diagnosis Tasks », *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Barcelone (Espagne), 2014.
- [2] Norme NF EN 10628, *Schémas de procédé pour l'industrie chimique et pétrochimique*, 2001.
- [3] Fay, A., et al., « Knowledge-based support of HAZOP studies using a CAEX plant model », *Inside Functional Safety*, vol. 2, p. 5–15, 2009.
- [4] Drath, R., et al., « Computer-aided design and implementation of interlocking control code », *IEEE International Symposium on Computer Aided Control Systems Design (CACSD)*, p. 2653–2658, 2006.
- [5] Barth, M., et al., « Object-oriented engineering data exchange as a base for automatic generation of simulation models », *35th Annual Conference of IEEE, Industrial Electronics (IECON'09)*, Porto (Portugal), 2009.
- [6] Yim, S. Y., et al., « Using process topology in plant-wide control loop performance assessment », *Computers and Chemical Engineering*, vol. 31, p. 86–99, 2006.
- [7] Norme CEI 62424, *Représentation de l'ingénierie de commande de processus – Demandes sous forme de diagrammes P&I et échange de données entre outils P&I et PCE-CAE*, août 2008.
- [8] Drath, R., *Datenaustausch in der Anlagenplanung mit AutomationML*, Springer-Verlag Berlin Heidelberg (Allemagne), 2010.
- [9] Holm, T., et al., « ISO 15926 vs. IEC 62424 – Comparison of Plant Structure Modeling Concepts », *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Cracovie (Pologne), 2012.