

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international



(43) Date de la publication internationale
3 mars 2011 (03.03.2011)

PCT

(10) Numéro de publication internationale
WO 2011/023876 A2

- (51) Classification internationale des brevets :
G06F 17/30 (2006.01) G06F 17/40 (2006.01)
G06Q 10/00 (2006.01)
- (21) Numéro de la demande internationale :
PCT/FR2010/051632
- (22) Date de dépôt international :
30 juillet 2010 (30.07.2010)
- (25) Langue de dépôt : français
- (26) Langue de publication : français
- (30) Données relatives à la priorité :
0955787 25 août 2009 (25.08.2009) FR
- (71) Déposant (pour tous les États désignés sauf US) :
CORAUD [FR/FR]; 53-55 Boulevard Romain Rolland,
F-92120 Montrouge (FR).
- (72) Inventeur; et
- (75) Inventeur/Déposant (pour US seulement) :
SOUBEYRAND, Philippe [FR/FR]; LA
PALOMBIERE, 22 Rue de Chéret, F-94000 Creteil (FR).
- (74) Mandataire : CABINET LHERMET LA BIGNE &
REMY; Emmanuel POTDEVIN, 11 boulevard de
Sébastopol, F-75001 Paris (FR).
- (81) États désignés (sauf indication contraire, pour tout titre
de protection nationale disponible) : AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Suite sur la page suivante]

(54) Title : METHOD FOR ORGANIZING VARIABLES IN A DATABASE

(54) Titre : PROCÉDE POUR ORGANISER DES VARIABLES DANS UNE BASE DE DONNEES

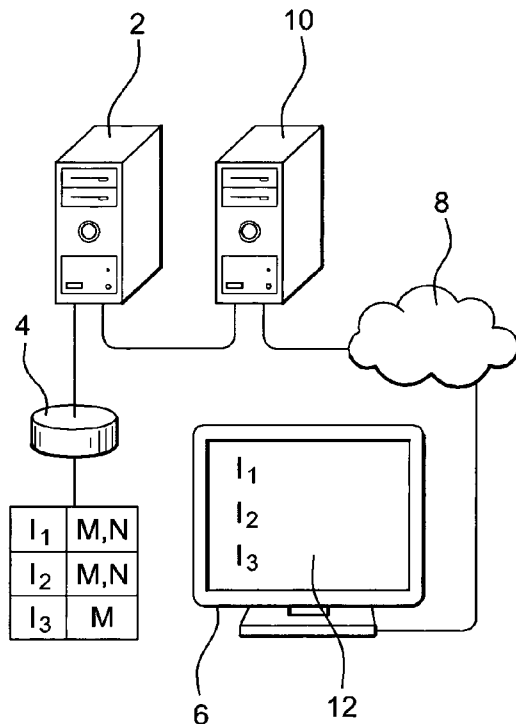


Fig. 30

(57) Abstract : The invention relates to a method for organizing a database (4) including variables, such as indicators (I_1, I_2, I_3) relating to an object assuming a plurality of states, wherein values associated with the respective states correspond to each variable. According to the method: automated means (2) identify correlations between at least some of the variables; the means (2) form groups, each consisting of one of the variables (I_3) and others (I_1, I_2) of the variables correlated with the latter and sufficient for deducing the latter; and the means (2) associate, in a database (4) and with each of the variables of the group, a datum (M) indicating at least one identity of the other variables of the group.

(57) Abrégé : Le procédé vise à organiser une base de données (4) comprenant des variables, telles que des indicateurs [I_1, I_2, I_3], relatives à un objet prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs. Dans le procédé : - des moyens automatisés (2) identifient des corrélations entre au moins certaines des variables; - les moyens (2) forment des groupes constitués chacun par une des variables (I_3) et d'autres [I_1, I_2] des variables corrélées à cette dernière et suffisant pour déduire cette dernière; et - les moyens (2) associent, dans une base de données (4), à chacune des variables du groupe, une donnée (M) indiquant au moins une identité des autres variables du groupe.



(84) États désignés (*sauf indication contraire, pour tout titre de protection régionale disponible*) : ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK,

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publiée :

— *sans rapport de recherche internationale, sera republiée dès réception de ce rapport (règle 48.2.g)*

Procédé pour organiser des variables dans une base de données

L'invention concerne les outils d'aide à la décision et l'organisation des variables dans une base de données.

On peut à titre d'exemple se référer au secteur des télécommunications. Dans ce dernier, l'ensemble de la production est mis en œuvre par le biais d'un système d'information qui gère à la fois les flux liés aux communications des clients (voix, SMS, MMS, Internet, etc...), les flux liés à la prospection et à la commercialisation, les flux liés au suivi financier, les flux liés au suivi de la qualité de service, les flux liés à la performance opérationnelle, etc.. Tous ces flux sont écoutés en temps réel au moyen d'agents informatiques (tels que des robots) qui collectent toutes sortes de données, avec des granularités plus ou moins fines selon les besoins des analyses (cela va généralement de la minute à l'heure).

Il existe généralement un agent par indicateur mesuré. Ces indicateurs vont de l'espace disque disponible sur une baie de disques durs, au débit informatique sur telle ou telle branche du réseau, en passant par les taux des charges des unités centrales des serveurs concernés, ou même plus simplement le nombre de connexions de clients sur tel ou tel serveur d'applications, etc... Ainsi, la personne en charge de l'analyse capacitaire (ou *capacity planning en anglais*) de l'entreprise dispose de l'ensemble des données dites d'apprentissage (l'histoire du système complexe observé) stocké au sein d'un entrepôt de données (en anglais *datawarehouse*).

Aussi, lorsque des personnes en charge des maîtrises d'œuvres ou d'ouvrages ont besoin d'un état du système d'information à un moment donné, l'analyste capacitaire est en mesure de leur en donner lecture. Ou même au-delà, lorsqu'une personne en charge de l'informatique décisionnelle (en anglais *business intelligence*) souhaite redéfinir des objectifs pour des activités au sein de l'entreprise (que l'on peut appeler objectifs métiers pour des chaînes métier), c'est cet analyste qui lui donne une lecture possible des indicateurs directement impactés par ces nouvelles prévisions. Ainsi, l'activité d'analyse capacitaire s'apparente à une surveillance permettant de donner une lecture de l'état de santé de l'entreprise, depuis les ventes réalisées en temps réel au sein des agences ou via Internet, jusqu'à l'ensemble des connexions réseau qu'il faut prévoir pour répondre à la demande.

En outre, l'analyse capacitaire permet d'établir des plans de contingence qui sont des outils d'aide à la décision : si une alerte est donnée quant au fonctionnement

d'un serveur d'application, il est souhaitable d'anticiper toutes les alertes à venir et de prendre les bonnes décisions afin d'éviter que la crise ne s'étende à l'ensemble de la production.

Mais jusqu'à présent, c'est principalement le retour d'expérience qui permet à l'analyste capacitaire d'envisager certaines conclusions.

En effet, le système d'information est en perpétuelle évolution et il est difficile pour l'analyste capacitaire d'anticiper les évolutions possibles (suppression ou ajout d'un serveur, diminution ou augmentation d'un ensemble de connexions, variation du nombre d'utilisateurs en agence, migration vers de nouvelles applications, etc.). De même, il lui est difficile de définir les nouveaux agents qui sont alors nécessaires à une lecture optimale de l'ensemble des indicateurs.

Plus généralement, on connaît des indicateurs appelés indicateurs clé de performance (ou KPI pour l'anglais *Key Performance Indicator*). Jusqu'à présent, le référentiel KPI est un outil dont la conception est réservée à l'informatique décisionnelle, qui procède en s'appuyant uniquement sur la nomenclature des métiers concernés. De son côté, l'analyste capacitaire doit garantir en permanence la tenue en charge du système complexe en fonction de la planification des objectifs métiers. Les experts de l'informatique décisionnelle font sans le savoir des choix qui influencent le système d'information. De leur côté, les analystes capacitaires se heurtent à la complexité du système d'information et appréhendent mal les chaînes métiers concernées depuis les indicateurs techniques observables jusqu'au KPI proprement dit. L'analyste tend à considérer que toute similitude remarquable entre deux indicateurs implique une relation de cause à effet, ce qui fausse considérablement les analyses capacitaires et les rapports de l'informatique décisionnelle. Il s'ensuit qu'on obtient des indicateurs souvent redondants ou surchargés au sein d'un système complexe devenu de facto compliqué, mal observé et coûteux, réduisant ainsi lourdement le retour sur investissement.

En outre, parmi les nombreux facteurs de dégradation de la qualité des données d'un entrepôt de données, il faut retenir la migration des données, la rupture des intégrités référentielles et les temps d'attente extrêmement long des clients du système complexe observé. Dans tous les cas, ce sont bien la stabilité et la fiabilité du gisement de données des indicateurs qui sont remises en cause. Et l'attribut commun de la qualité des données qui est remis en cause reste l'intégrité du référentiel d'indicateurs afin de ne pas effacer ou de modifier par inadvertance les données d'une table en relation avec une autre.

Un but de l'invention est donc d'améliorer la qualité des décisions reposant sur des variables telles que des indicateurs.

A cet effet, on prévoit, selon l'invention, un procédé pour organiser une base de données comprenant des variables, telles que des indicateurs, relatives à un objet
5 prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs, procédé dans lequel :

- des moyens automatisés identifient des corrélations entre au moins certaines des variables ;

10 - les moyens forment des groupes constitués chacun par une des variables et d'autres des variables corrélées à cette dernière et suffisant pour déduire cette dernière ; et

- les moyens associent, dans une base de données, à chacune des variables du groupe, une donnée indiquant au moins une identité des autres variables du groupe.

15 Ainsi, l'invention procède en s'appuyant sur les corrélations réellement observables au sein du gisement de données des variables relatives au système complexe observé. Elle assure la corrélation et la codification significative des variables clé au sein de l'analyse capacitaire et permet d'établir une véritable cartographie des variables entre elles, tout en consolidant l'intégrité des données au sein de l'entrepôt de données (authenticité et traçabilité). Le procédé permet
20 d'organiser des variables discrètes et chaotiques en une arborescence montrant les relations de cause à effet probables, depuis les variables dites élémentaires, jusqu'aux variables clé. L'invention permet une meilleure compréhension du système d'information en temps réel. Elle rend possible une automatisation des plans de
25 contingence, des opérations de surveillance et de la gestion des alertes. L'invention permet une lecture rapide des impacts liés aux prévisions (côté informatique décisionnelle) comme aux crises (côté maîtrise d'œuvre ou d'ouvrage). Mieux qu'une étude d'impact, la base obtenue avec l'invention peut-être mise à jour régulièrement afin de prendre en compte les dernières évolutions au sein du système d'information.

30 Avantageusement, les moyens construisent un graphe, tel qu'un réseau de Petri, du groupe, par exemple au moyen d'un algorithme montant.

Ainsi le graphe encapsule l'identité de la variable clé et le groupe auquel il appartient.

Avantageusement, les moyens codent le graphe pour produire la donnée.

35 Ce code permet de retrouver rapidement et facilement le graphe correspondant, et donc les identités des variables et leurs relations. Cette codification significative des variables renforce la qualité des données au sein d'un entrepôt de données, en

termes d'intégrité d'abord (authenticité et traçabilité), mais aussi en termes de complétude. Cette forme de représentation comme graphe, combinée à la codification significative, autorise un traitement ultérieur par exemple d'hypersphérisation ou d'iconographie, ainsi qu'une modélisation de type file d'attente.

Dans un mode de réalisation :

- on sélectionne une des variables et on place un vecteur correspondant à la variable;
- on sélectionne une des variables qui lui est le plus étroitement corrélée et on place un vecteur correspondant à cette variable à une position représentant cette corrélation ;

- puis, autant de fois que nécessaire, on sélectionne une des variables qui est le plus étroitement corrélée à celles correspondant à un vecteur déjà positionné et on place un vecteur correspondant à cette variable à une position représentant cette corrélation,

les vecteurs étant positionnés avec leur origine au centre d'une sphère représentant l'objet et leur extrémité sur la sphère.

Ce mode de représentation permet d'évaluer le degré de pertinence à l'égard de l'objet des variables considérées et la complétude de l'outil de supervision vis-à-vis du système complexe observé.

On prévoit aussi selon l'invention un procédé pour exploiter des variables, telles que des indicateurs, relatives à un objet prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs, procédé dans lequel ;

- des moyens automatisés obtiennent dans une base de données au moins une donnée indiquant une identité de variables corrélées à une variable prédéterminée et formant avec elle un groupe tel que les variables du groupe, sauf une, suffisent pour déduire cette dernière, les moyens identifiant ainsi les variables corrélées ; et

- les moyens exploitent les identités des variables identifiées, par exemple les transmettent sur un réseau de télécommunication.

De préférence, les moyens identifient à partir de la donnée des relations de cause à effet entre les variables du groupe.

Avantageusement, les moyens commandent la mention sur une page web des identités des variables identifiées.

On prévoit également selon l'invention un programme d'ordinateur qui comprend des instructions de code apte à commander la mise en œuvre d'un procédé selon l'invention lorsqu'il est exécuté sur un ordinateur.

On prévoit aussi la mise à disposition de ce programme sur un réseau de télécommunication en vue de son téléchargement, ainsi qu'un support d'enregistrement de données comprenant ce programme sous forme enregistrée.

On prévoit aussi selon l'invention un dispositif pour organiser une base de données comprenant des variables, telles que des indicateurs, relatives à un objet
5 prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs, le dispositif comprenant des moyens aptes :

- à identifier des corrélations entre au moins certaines des variables ;

- à former des groupes constitués chacun par une des variables et d'autres des
10 variables corrélées à cette dernière et suffisant pour déduire cette dernière ; et

- à associer, dans une base de données, à chacune des variables du groupe, une donnée indiquant au moins une identité des autres variables du groupe.

On prévoit également selon l'invention un dispositif d'exploitation de variables, telles que des indicateurs, relatives à un objet prenant plusieurs états, à chaque
15 variable correspondant des valeurs associées aux états respectifs, le dispositif, éventuellement conforme au dispositif précité, comprenant des moyens aptes :

- à obtenir dans une base de données au moins une donnée indiquant une identité de variables corrélées à une variable prédéterminée et formant avec elle un
20 groupe tel que les variables du groupe, sauf une, suffisent pour déduire cette dernière, les moyens étant aptes ainsi à identifier les variables corrélées ; et

- à exploiter les identités des variables identifiées, par exemple à les transmettre sur un réseau de télécommunication.

On prévoit par ailleurs selon l'invention une base de données qui comprend :

- des variables relatives à un objet prenant plusieurs états, à chaque variable
25 correspondant des valeurs associées aux états respectifs;

- au moins une donnée en relation avec au moins une des variables et indiquant une identité de certaines des variables corrélées avec elle et formant avec elle un
groupe tel que les variables du groupe, sauf une, suffisent pour déduire cette dernière.

30 On peut employer l'invention auprès d'entités ayant des besoins dans les domaines de l'informatique décisionnelle ou de l'analyse capacitaire. Tous les secteurs industriels sont concernés par cette invention : nucléaire, transport, pharmaceutique, cosmétique, automobile, équipementiers, énergie, financier, assurance, etc... En effet, quel que soit le domaine d'application, peu importent les
35 noms qui sont donnés aux variables, dès lors que les exigences propres à la

technique sont respectées et qu'une lecture en temps réel est possible. L'invention peut ainsi s'appliquer par exemple dans les domaines suivants:

- nucléaire : température intérieure d'une boîte à gants, température de la matière chaude en cours d'usinage, débit du lubrifiant, épaisseur des gants de manipulation, temps d'ouverture des hublots, temps d'exposition des techniciens, *etc*

- pharmaceutique : rendement d'un flot automatisé de conditionnement, décompte des gélules non conformes, décompte des substrats non conformes, vitesse de mise en place sous vide, décompte des substrats détériorés suite à la mise en place, *etc...*

- énergie : débit d'un fleuve, température de l'eau, débit du circuit de refroidissement, puissance du générateur, température du cœur, demande en énergie, consommation propre à la centrale, quantité d'énergie disponible, *etc...*

- économie : indices financiers et boursiers tels que le CAC 40, volume d'actions échangées, taux directeurs, taux interbancaires, volume de crédit, indice de volatilité, évolution du PIB, évolution du chômage, moral des entreprises, moral des consommateurs, *etc...*

L'invention sera mieux comprise à la lecture de la description qui va suivre, donnée uniquement à titre d'exemple et faite en se référant aux dessins dans lesquels :

- la figure 1 est un organigramme illustrant la correspondance entre un langage formel et une famille de métiers ;

- la figure 2 présente des courbes illustrant la liaison par superposition entre deux indicateurs ;

- les figures 3 et 4 sont des réseaux de Petri d'un caractère redondant et d'une liaison d'indicateurs ;

- les figures 5 et 6 sont respectivement une matrice de connexités et l'empilement des indicateurs y relatifs ;

- les figures 7 et 8 illustrent les mêmes éléments dans le cas d'un empilement parallèle ;

- la figure 9 est le réseau de Petri d'un empilement linéaire des indicateurs élémentaires au sein d'une chaîne ;

- les figures 10 et 11 montrent des matrices dans lesquelles respectivement les indicateurs sont redondants ou encore surchargent l'un des indicateurs ;

- les figures 12 à 14 sont des matrices illustrant respectivement l'initialisation de la matrice descriptive dans le procédé de l'invention, la définition de la zone élémentaire et enfin la forme générale en escalier ;

- les figures 15 à 21 sont des matrices et des réseaux de Petri illustrant respectivement une redondance simple, un doublon, une redondance multiple, une liaison d'indicateurs, une surcharge d'indicateurs, une surcharge redondante et enfin une situation d'interprétation générale ;
- 5 - la figure 22 illustre un exemple d'un réseau de Petri parallèle ;
- la figure 23 montre une matrice descriptive et les connexités dans le cas du réseau de Petri parallèle ;
- la figure 24 illustre un exemple du réseau de Petri linéaire ;
- la figure 25 présente la matrice descriptive et les connexités dans le cas du réseau
10 de Petri linéaire ;
- la figure 26 illustre un graphe des liaisons d'une chaîne KPI ;
- la figure 27 montre l'empilement des indicateurs élémentaires au sein de la chaîne KPI ;
- la figure 28 montre la matrice descriptive et les connexités au sein de cette chaîne ;
- 15 - la figure 29 montre un réseau de Petri dans un exemple de réalisation; et
- la figure 30 est un schéma illustrant des moyens de mise en œuvre de l'invention ;
- les figures 31 à 36 et 38 à 42 sont des diagrammes illustrant la construction de la représentation en trois dimension du système de supervision des variables ;
- la figure 37 illustre un principe de cette représentation ; et
- 20 - les figures 43 et 44 sont deux exemples de représentations en résultant.

Nous allons présenter ci-après des exemples de mise en œuvre de l'invention dans lesquels les variables sont des indicateurs, l'invention n'étant pas cependant réduite à ce cas.

Il s'agit d'un procédé de codification significative des indicateurs clés de performance ou KPI (pour l'anglais *Key Performance Indicators*). Il met en œuvre en
25 l'espèce :

- des calculs de corrélation des indicateurs d'un système complexe observé (matrice de variances-covariances, matrice de corrélations mathématiques et matrice de corrélations croisées);
- 30 - la construction et l'interprétation de matrices dites descriptives et de connexités, et
- la représentation de l'ensemble des chaînes KPI modélisées sous la forme de graphes tels que des réseaux de Petri.

Le procédé met en oeuvre une grammaire des graphes dédiée qui permet
35 d'automatiser la modélisation des réseaux de Petri à partir de la simple relecture de la codification significative des KPI ainsi obtenue. Cette codification constitue la

nouvelle clé d'entrée d'un entrepôt de données (gisement de données des indicateurs techniques et métiers propre au système observé) et facilite la mise en place d'une base de donnée moyennant par exemple une technologie des groupes assistée par ordinateur (ou TGAO), dédiée à l'activité d'analyse capacitaire.

5 Nous allons expliquer d'abord comment définir, construire et utiliser la base de données propre à l'entrepôt de données.

On utilise une classification des chaînes métiers étudiées qui désignent ici des groupes de variables associés à une même activité au sein du système observé (vente, production, qualité etc.). Il s'agit de retrouver rapidement leurs codes grâce à l'usage d'un thésaurus dans la base. Pour cela, nous construisons une codification significative, voir figurative, de ces chaînes métiers, puis nous associons au résultat ainsi obtenu une codification chronologique, le tout constituant la nouvelle clef d'entrée de l'entrepôt de données.

15 Un thésaurus est un dictionnaire hiérarchisé, un vocabulaire normalisé sur la base de termes génériques et de termes spécifiques à un domaine. Il ne fournit qu'accessoirement des définitions, les relations des termes et leur choix l'emportant sur les significations. Toutefois, lorsqu'un thésaurus s'appuie sur une codification significative, la codification d'un objet n'est plus seulement la référence de celui-ci, mais porte en plus une sémantique précisant le cas d'emploi. Un thésaurus est finalement une arborescence d'ensembles ordonnés de références, une classification n'étant que le simple fait de mettre en relation des objets avec ces références. Dans le cadre de l'activité d'analyse capacitaire, ces objets sont ici les chaînes métiers.

20 On considère ici qu'une codification est un ensemble de règles définissant une correspondance biunivoque (bijection) entre des informations et leur représentation par des caractères, des symboles ou des éléments de signal. Il peut s'agir d'une codification dite soit significative soit chronologique. Pour toute chaîne métier ainsi référencée au sein d'une classification propre à l'analyse capacitaire, il existe une référence unique permettant de l'authentifier et de la tracer.

30 Une codification figurative est une méthode mnémotechnique permettant l'accès aux données et aux informations propres à un objet considéré. De préférence, cette codification ne s'applique qu'à des objets immuables dans le temps. Dans tous les autres cas, la mise en place du thésaurus n'en devient que plus complexe, perdant aussi le bénéfice de la cohérence interne de ce dernier.

35 Une grammaire est un formalisme permettant de définir une syntaxe et donc un langage formel, c'est-à-dire un ensemble de mots sur un alphabet donné. En d'autres termes, une grammaire fixe les règles permettant d'écrire un code non arbitraire. Il

existe une similitude entre la génération des phrases à partir des symboles (mots) respectant une sémantique liée à une langue donnée et la génération des chaînes métiers à partir des indicateurs métiers ou techniques, en considérant les relations de cause à effet. La Fig. 1 illustre ainsi cette correspondance entre un langage formel et une famille métiers.

Nous rappelons ci-après quelques outils de corrélation.

Soit un vecteur aléatoire \vec{X} à N indicateurs temporels discrets et chaotiques X_j présentant des pas de collecte différents au sein de l'entrepôt de données. La première étape consiste donc en une interpolation pertinente des indicateurs les plus pauvres de sorte que chaque indicateur puisse disposer de n réalisations au sein de la même plage de temps:

$$\vec{X} = \begin{pmatrix} X_1 = (x_{11} \dots x_{1n}) \\ \vdots \\ X_N = (x_{N1} \dots x_{Nn}) \end{pmatrix}$$

Une fois cette préparation réalisée, les calculs concernent essentiellement les éléments suivants:

- la matrice de variances-covariances :

$$\Sigma_{\vec{X}} = \begin{bmatrix} \sigma_{X_1}^2 & & & & \dots & \\ & \ddots & & & & \\ & & \sigma_{X_k}^2 & & & \\ & & & \sigma_{X_j X_i} & & \\ & \sigma_{X_i X_j} & & & \ddots & \\ \dots & & & & & \sigma_{X_N}^2 \end{bmatrix}_{N,N}$$

sachant que l'ensemble des calculs s'appuie sur la loi de Bravais-Pearson où

$$\sigma_{xy} = \text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \overline{x \cdot y} - \bar{x} \cdot \bar{y}$$

avec $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ et où $\sigma_x = \sqrt{\text{var}(x)}$ avec

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$$

- la matrice de corrélations :

$$\Pi_{\bar{X}} = \begin{bmatrix} 1 & & & \dots \\ & \ddots & & \\ & & 1 & \frac{\sigma_{X_j X_i}}{\sigma_{X_j} \sigma_{X_i}} \\ & \frac{\sigma_{X_i X_j}}{\sigma_{X_i} \sigma_{X_j}} & & \ddots \\ \dots & & & & 1 \end{bmatrix}_{N.N}$$

- la matrice de corrélations croisées (il est intéressant d'avoir les retards entre les indicateurs) :

$$\Gamma_{\bar{X}} = \begin{bmatrix} 0 & & & \dots \\ & \ddots & & \\ & & 0 & \tau_{X_j X_i} \\ & \tau_{X_i X_j} & & \ddots \\ \dots & & & & 0 \end{bmatrix}_{N.N}$$

5 sachant que l'ensemble des calculs s'appuie sur la méthode de corrélation croisée où $\Gamma_{XY}(\tau) = E(X(t)Y(t-\tau))$ E désignant l'espérance mathématique et

où $\Gamma_{XY}(\tau) = \int_{-\infty}^{+\infty} X(t)Y^*(t-\tau)dt$ avec Y^* désignant la transformée de la

fonction Y telle que $Y^*(t-\tau) = Y(t-\tau)\delta_{T_e}(t-\tau)$ et $\delta_{T_e}(t-\tau) = \sum_{k=-\infty}^{+\infty} \delta(t-\tau-kT_e)$,

δ_{T_e} étant le peigne de Dirac de période d'échantillonnage T_e

10 Nous détaillons plus loin le procédé de mise en œuvre de ces calculs permettant de proposer les relations de cause à effet au sein d'une chaîne métier, et d'en déduire une certaine forme de représentation, puis de codification significative.

Lors de la mise en place d'un entrepôt de données, on démarre par l'inventaire complet des composants constituant le modèle générique et unique, afin de donner
15 une représentation fidèle de l'objet étudié. La construction de la base passe par les étapes suivantes :

- inventaire des données des chaînes métiers et des indicateurs attachés ;
- construction de la matrice des corrélations entre indicateurs deux à deux ;
- codification des chaînes métiers depuis les indicateurs élémentaires identifiés ;
- 20 - classification des chaînes métiers codifiées ;
- relecture du thésaurus ;
- validation de la grammaire des graphes utilisée ;
- modélisation générique des chaînes métiers.

Cette construction se fait selon une lecture séquentielle de ces étapes. Quant à son utilisation, nous procédons de la manière suivante :

- depuis la classification au sein du thésaurus, sélection du code d'une chaîne métier ;

5 - depuis la grammaire des graphes retenue, lecture de l'ensemble des relations de cause à effet au sein de cette chaîne métier ;

- depuis le modèle de cette chaîne métier, hypersphérisation pour une représentation à la fois intuitive et cognitive.

10 Nous allons voir dans la deuxième partie comment nous pouvons automatiser l'écriture des chaînes métiers depuis les calculs de corrélation, et ainsi construire la base.

15 Un processus de décision dépend directement de la qualité des données stockées au sein de l'entrepôt. C'est pourquoi il est préférable de réaliser un inventaire des attributs de la qualité des données propres aux indicateurs. Pour chacun de ces attributs, nous établissons une définition, bâtissons une métrique, mesurons le pointage métrique et suivons la progression des mesures en temps réel. Ces attributs sont ici les suivants: la consistance, la complétude, l'accessibilité, l'intégrité, la précision, la validité et l'opportunité.

20 En gestion de base de données, une intégrité référentielle est un jeu de règles informatiques qui empêche d'effacer ou de modifier par inadvertance les données d'une table en relation avec une autre. De manière générale, l'intégrité désigne l'état des données qui, lors d'un traitement, d'une conservation ou d'une transmission, ne subissent aucune altération ou destruction, et conservent un format permettant leur réutilisation. La notion d'intégrité des données encapsule deux autres attributs de la

25 qualité des données :

- authenticité : chaque donnée dispose d'une identité (preuve) unique et vérifiable ;

- traçabilité : chaque donnée appartient à une chaîne (parcours) facilement identifiable.

30 Concernant l'authenticité, cet attribut est généralement traité par le biais d'algorithmes de type MAC, *Message Authentication Code* connus en eux-mêmes. Concernant la traçabilité, cet attribut nous amène à la généralisation d'une codification universelle, qui doit être lue et comprise par tous, notamment ici par l'entrepôt de données.

35 Nous allons maintenant présenter comment on obtient les graphes et la codification significative.

Observons le cas particulier de trois indicateurs dont les données sont supposées être exprimées dans la même unité au sein du système international d'unités (SI).

5 Soit une chaîne métier \mathcal{M} formant un groupe de trois indicateurs tel que $\mathcal{M} = \{I_i, I_j, I_k\}$. La figure 2 illustre une liaison par superposition entre deux courbes d'indicateurs. Dans le couple $(r_{p_{ik}}, \varphi_{ik})$, $r_{p_{ik}}$ désigne le coefficient de corrélation mathématique entre I_k et I_i , φ_{ik} désigne le retard croisé de I_k par rapport à I_i . Dans le couple $(r_{p_{jk}}, \varphi_{jk})$, $r_{p_{jk}}$ désigne le coefficient de corrélation mathématique entre I_k et I_j , φ_{jk} désigne le retard croisé de I_k par rapport à I_j .

10 Au sens graphique, nous observons une liaison par superposition entre les cours des indicateurs I_i et I_j . Nous notons cette liaison L_{ij} ou L_{ji} , puisque nous pouvons considérer I_i sous I_j ou I_j sous I_i . Quels que soient les retards observés par rapport aux cours d'indicateurs sources, l'indicateur résultat I_k est toujours obtenu par superposition des deux autres tout en tenant compte de leurs niveaux de contribution
15 respectifs. En outre, on distingue le niveau de contribution et le coefficient de corrélation qui n'expriment pas la même chose. Au sens de la liaison, I_i et I_j sont les indicateurs élémentaires. I_k devient l'indicateur induit.

20 La recherche de l'ensemble des graphes permettant d'aboutir à la représentation de toute chaîne métier nécessite d'abord la définition d'un modèle unique permettant de représenter toutes les chaînes métiers possibles. On donne les définitions suivantes :

- indicateur élémentaire : tous les indicateurs non corrélés entre eux deux à deux au sein d'une chaîne donnée, sont des indicateurs élémentaires ;

25 - indicateur induit : tout indicateur qui n'est pas un indicateur élémentaire est induit ;

- indicateur KPI : un indicateur induit devient un indicateur KPI au sens du groupe si et seulement si il ne permet pas de déduire d'autres des indicateurs. Un indicateur KPI se trouve toujours placé en bout de chaîne à l'issue du traitement de corrélation ;

30 - chaîne KPI : tout groupe, ici une chaîne, constitué d'indicateurs élémentaires, induits et d'un KPI dont les enchainements sont tous connus et issus des calculs de corrélation. Il peut y avoir plusieurs chaînes KPI au sein d'une seule chaîne métier ;

- liaison d'indicateurs : une liaison entre deux indicateurs élémentaires ou induits se traduit toujours par la superposition des cours de ces indicateurs, tout en

tenant compte de leurs niveaux de contribution respectifs, afin de créer un nouvel indicateur induit au sein de la chaîne métier observée. Au-delà de deux indicateurs, nous disons que la liaison d'indicateurs est surchargée ;

- 5 - caractère redondant : il apparaît si un indicateur induit ne dépend que d'un indicateur élémentaire ou induit;
- coefficient SI : coefficient de transformation d'un indicateur élémentaire, induit ou KPI pour l'exprimer dans les unités du système international ;
- coefficient de contribution : tout niveau de contribution, exprimé dans le SI, d'un indicateur sur un autre. Exemple : une variation de la valeur de I_k de 10 unités engendre une variation de la valeur de I_l de l'ordre de -1000 de ces mêmes unités ;
- 10 - corrélation mathématique : le coefficient de corrélation mathématique entre deux indicateurs est ici reconnu valable dès que sa valeur absolue est supérieure ou égale à $\frac{\sqrt{3}}{2} \approx 0,866$; dès lors qu'elle tend vers 0, les deux indicateurs sont dits indépendants ;
- 15 - covariance d'indicateurs : la covariance d'indicateurs est reconnue valable dès que sa valeur absolue converge vers 0. Dans ce cas seulement, nous pouvons affirmer que ces indicateurs sont indépendants et qu'ils ne sont jamais corrélés. Dans ce calcul, le seuil de convergence vers 0 doit être apprécié à l'issue de l'expérience ;
- retard croisé : deux indicateurs corrélés entre eux ne sont vraisemblablement pas simultanés. Le retard croisé permet de quantifier ce décalage de réalisation dans le temps. Nous parlons aussi de déphasage dans le temps ;
- 20 - relation de cause à effet exprimée : le retour d'expérience peut s'avérer pertinent lorsque le calcul seul ne permet pas de conclure ;
- entropie de Shannon : lorsqu'un indicateur devient instable au sens de la distribution, son entropie de Shannon devient minimale et se rapproche brutalement de 0.
- 25

Soit un groupe tel qu'une chaîne métier \mathcal{M} constituée de M indicateurs dont un nombre nb d'entre eux sont des indicateurs élémentaires.

Soit un groupe tel qu'une chaîne KPI \mathcal{M}_{kpi} constituée de N indicateurs pris parmi ceux de la chaîne métier \mathcal{M} et dont m d'entre eux sont des indicateurs élémentaires.

Nous devons respecter les inégalités suivantes: $m \leq nb < M$ et $m < N \leq M$.

Nous pouvons noter les indicateurs $I_i \forall i \in \{1 \dots M\}$.

Nous pouvons noter la chaîne métier $\mathcal{M} = \{I_1, \dots, I_{nb}, \dots, I_M\}$;

Soit \bar{X}_M le vecteur constitué des M indicateurs présents dans la chaîne métier \mathcal{M} ;

Soit A_M la matrice descriptive de \mathcal{M} telle que:

$$- A_M \in \Pi_{M \times M} / \Pi = \{0,1\} ;$$

- 5 - $\forall i, j \in \{1 \dots M\}$, nous associons à toute corrélation identifiée valable et notée $r_{p_{ij}}$ le nombre $\pi_{ij} = 1$, ou 0 s'il n'y a pas de corrélation valable sachant que $\pi_{ii} = 1$, $\forall i \in \{1 \dots M\}$;

- $A_M = [a_{ij}] / a_{ij} = \pi_{ij} = \pi_{ji} = a_{ji}$; A_M est une matrice carrée creuse et symétrique et dont la trace est toujours égale à M .

- 10 Soit A_{nb} la matrice de connexité de \mathcal{M} telle que:

$$- A_{nb} \in \nabla_{nb \times nb} / \nabla = \{0,1\} ;$$

- $\forall i, j \in \{1 \dots nb\}$, nous associons à toute liaison identifiée et notée L_{ij} le nombre $\delta_{ij} = 1$, ou 0 s'il n'y a pas de liaison sachant que $\delta_{ii} = 0$, $\forall i \in \{1 \dots nb\}$;

- $A_{nb} = [a_{ij}] / a_{ij} = \delta_{ij} = \delta_{ji} = a_{ji}$; A_{nb} est une matrice carrée creuse et symétrique et dont la diagonale est toujours nulle.

15

Nous pouvons noter la chaîne KPI $\mathcal{M}_{kpi} = \{I_1, \dots, I_m, \dots, I_N\}$.

Soit \bar{X}_N le vecteur constitué des N indicateurs présents dans la chaîne KPI \mathcal{M}_{kpi} .

Soit A_N la matrice descriptive de \mathcal{M}_{kpi} telle que:

$$- A_N \in \Pi_{N \times N} / \Pi = \{0,1\} ;$$

20

- $\forall i, j \in \{1 \dots N\}$, nous associons à toute corrélation identifiée valable et notée $r_{p_{ij}}$ le nombre $\pi_{ij} = 1$, ou 0 s'il n'y a pas de corrélation valable sachant que $\pi_{ii} = 1$, $\forall i \in \{1 \dots N\}$;

- $A_N = [a_{ij}] / a_{ij} = \pi_{ij} = \pi_{ji} = a_{ji}$; A_N est une matrice carrée creuse symétrique et dont la trace est toujours égale à N .

25

Soit A_m la matrice de connexité de \mathcal{M}_{kpi} telle que:

$$- A_m \in \nabla_{m \times m} / \nabla = \{0,1\} ;$$

- $\forall i, j \in \{1 \dots m\}$, nous associons à toute liaison identifiée et notée L_{ij} le nombre $\delta_{ij} = 1$, ou 0 s'il n'y a pas de liaison sachant que $\delta_{ii} = 0, \forall i \in \{1 \dots m\}$;

- $A_m = [a_{ij}] / a_{ij} = \delta_{ij} = \delta_{ji} = a_{ji}$; A_m est une matrice carrée creuse symétrique et dont la diagonale est toujours nulle.

5 Soient les représentations suivantes sous forme de réseau de Petri:

- caractère redondant : La figure 3 illustre le réseau de Petri d'un caractère redondant.

avec $r_{p_{ij}} \in [-1, 1]$ et $\varphi_{ij} \geq 0$;

$$I_j(t) \equiv \alpha_{ij} I_i(t - \varphi_{ij})$$

10 Cette formule donne la relation entre deux indicateurs redondants avec:

- $\alpha_{ij} = \frac{\alpha_i}{\alpha_j} \cdot c_{ij}$ le coefficient de contribution de I_i sur I_j

- α_i et α_j les coefficients SI des indicateurs observés et

- c_{ij} le coefficient de contribution exprimé dans le SI de I_i sur I_j .

- liaison d'indicateurs : La figure 4 illustre le réseau de Petri d'une liaison

15 d'indicateurs. Avec $r_{p_{uw}} \in [-1, 1], r_{p_{vw}} \in [-1, 1], \varphi_{uw} \geq 0$ et $\varphi_{vw} \geq 0$

$$I_w(t) \equiv \beta_{uw} I_u(t - \varphi_{uw}) + \gamma_{vw} I_v(t - \varphi_{vw})$$

avec:

- $\beta_{uw} = \frac{\beta_u}{\beta_w} \cdot c_{uw}$ le coefficient de contribution de I_u sur I_w ,

- β_u et β_w les coefficients SI de ces indicateurs observés et

20 - c_{uw} le coefficient de contribution exprimé dans le SI de I_u sur I_w ;

- $\gamma_{vw} = \frac{\gamma_v}{\gamma_w} \cdot c_{vw}$ le coefficient de contribution de I_v sur I_w ,

- γ_v et γ_w les coefficients SI de ces indicateurs observés et

- c_{vw} le coefficient de contribution exprimé dans le SI de I_v sur I_w .

25 Nous utilisons une forme de représentation simple et accessible de type réseau de Petri qui, lorsqu'elle est temporisée et/ou colorée, autorise toujours toute démarche ultérieure et complémentaire d'hypersphérisation ou d'analyse sous forme de files d'attentes.

Le niveau de complexité du problème lié aux calculs des corrélations croît avec le nombre d'indicateurs constituant la chaîne métier observée. En effet, le nombre de liaisons par superposition au sein de la chaîne métier étudiée : $\mathcal{M} = \{I_1, \dots, I_{nb}, \dots, I_M\}$, dépend directement du nombre nb d'indicateurs élémentaires présents au sein de la chaîne métier observée (tout KPI mis à part). Le nombre nb est toujours supérieur ou égal à m . C'est pourquoi, nous nous arrêtons d'abord au cas particulier d'une chaîne métier ayant un seul indicateur KPI.

La construction de cette chaîne revient à empiler linéairement (tout coefficient de contribution mis à part) l'ensemble des indicateurs élémentaires par le biais de superpositions successives. Dans ce cas, les chaînes métier et KPI sont strictement identiques ; $N = M$ et $nb = m$. Lors d'un empilement linéaire, la matrice de connexité de $\mathcal{M}_{kpi} = \{I_1, \dots, I_m, \dots, I_N\}$ est telle qu'illustré à la figure 5. La figure 6 illustre l'empilement des indicateurs élémentaires au sein de \mathcal{M}_{kpi} et montre ce qui se passe au sein de la chaîne KPI étudiée. De ces deux représentations, nous déduisons que le nombre de liaisons par superposition présent dans \mathcal{M}_{kpi} est $m - 1$.

Il se peut lors du calcul des corrélations qu'une chaîne métier fasse ressortir non pas un KPI, mais plusieurs. Dans ce cas, chaque chaîne KPI doit être traitée distinctement. Dans le cas d'une chaîne ayant deux KPI, l'empilement est nécessairement parallèle, comme le montre la figure 8 qui illustre l'empilement des nb indicateurs élémentaires au sein de \mathcal{M} s'il y a deux KPI. Dans cet exemple, la matrice de connexité de \mathcal{M} prend la forme illustrée à la figure 7. Nous en déduisons que la forme générale de la matrice de connexité permet de retrouver le nombre de KPI réellement présents au sein de la chaîne métier \mathcal{M} . Dans le cas présenté ci-dessus, la chaîne métier dispose donc de deux chaînes KPI : \mathcal{M}_{kpi1} et \mathcal{M}_{kpi2} .

De façon générale, chaque chaîne KPI devant être traitée distinctement, l'empilement retenu est toujours linéaire. Ainsi, le nombre de liaisons par superposition au sein de la chaîne KPI \mathcal{M}_{kpi} est $m - 1$. Nous décrivons donc la chaîne KPI $\mathcal{M}_{kpi} = \{I_1, \dots, I_m, \dots, I_N\}$ comme sur la figure 9 qui illustre le réseau de Petri de l'empilement linéaire des indicateurs élémentaires au sein de \mathcal{M}_{kpi} .

Soit $G_{\mathcal{N}}$ l'ensemble des réseaux de Petri possibles. Dans ce cas, le cardinal de $G_{\mathcal{N}}$ est tel que

$$\text{card } G_N = C_m^2 P_{m-2} = \left(\frac{A_m^2}{2!} \right) P_{m-2} = \frac{m!}{2!(m-2)!} (m-2)! = \frac{m!}{2}$$

avec :

- C_m^2 , le nombre de combinaisons de 2 indicateurs élémentaires pris parmi m ;
- P_{m-2} , le nombre de permutations des $m-2$ indicateurs élémentaires restants ;

5 et

- A_m^2 , le nombre d'arrangements de 2 indicateurs élémentaires pris parmi m .

La forme générale du réseau de Petri nous permet de préciser l'un des attributs de la qualité des données pour la chaîne étudiée, à savoir la complétude (toutes les données nécessaires sont disponibles dans la base). Ainsi, pour une chaîne KPI étudiée \mathcal{M}_{kpi} , dont la forme générale du réseau de Petri est elle aussi linéaire, la complétude est optimale si et seulement si le nombre d'indicateurs élémentaires est tel que :

10

$$\forall N \text{ impair } \geq 1 : m = \frac{N-1}{2} + 1$$

$$\forall N \text{ pair } \geq 2 : m = \frac{N}{2}$$

15

Ainsi, le card G_{20} à complétude optimale :

$$\text{card } G_{20} = \frac{10!}{2} = 1814400 = \text{card } G_{19}$$

De même, le card G_{100} à complétude optimale s'écrit:

$$\text{card } G_{100} = \frac{50!}{2} = 1,5207046600856689021806304083032 \cdot 10^{64} = \text{card } G_{99}$$

20

Au vu de la complexité des calculs matriciels qui peuvent être exécutés lors des calculs de corrélation, pour arrêter le choix définitif du réseau de Petri décrivant exactement la chaîne KPI $\mathcal{M}_{kpi} = \{I_1, \dots, I_m, \dots, I_N\}$, nous retenons les deux critères suivants: la complétude des données et l'ordonnancement définitif des indicateurs selon l'ensemble des retards croisés calculés.

25

Le nombre d'indicateurs élémentaires réellement obtenus dans une chaîne KPI nous renseigne sur la pertinence des choix qui sont faits.

Supposons que dans un premier cas, la chaîne KPI est trop écoutée. Les indicateurs tendent alors à devenir redondants et la matrice descriptive « s'horizontalise ». Nous pouvons réaliser des économies de mesures et établir des

relations mathématiques afin de déduire par le calcul l'ensemble des indicateurs redondants depuis l'indicateur réalisé au plus tôt. La matrice est illustrée à la figure 10. Dans cet exemple, les indicateurs I_j , I_k et I_l sont redondants.

Supposons que dans un second cas, la chaîne KPI n'est pas assez écoutée. Les indicateurs tendent alors à surcharger certaines liaisons d'indicateurs et la matrice descriptive « se verticalise ». Nous pouvons compléter la chaîne métier initiale et ajouter de nouveaux indicateurs induits pertinents afin de décomposer linéairement toute liaison surchargée depuis la liaison réalisée au plus tôt. La matrice est illustrée à la figure 11. Dans cet exemple, les indicateurs I_j , I_k et I_l surchargent I_i .

Dans les deux cas, c'est bien l'attribut de la qualité des données, la complétude, qui doit être renforcé, sans quoi le retour d'expérience peut très tôt s'avérer nécessaire par l'usage de relations de cause à effet exprimées par l'expert, ce que nous voulons éviter. De préférence, les calculs de corrélation ne doivent pas être orientés par les avis des experts.

Nous présentons maintenant la construction des matrices descriptives et de connexités. Voici quelle méthode de corrélation nous pouvons retenir pour expliquer une chaîne métier.

Soit une chaîne métier $\mathcal{M} = \{I_1, \dots, I_{nb}, \dots, I_M\}$.

En tenant compte de la granularité la plus grossière, nous échantillons l'ensemble des indicateurs de cette chaîne afin de ne retenir que n réalisations ou observations. Dans certains cas, il est possible d'interpoler les réalisations manquantes sur des indicateurs dont la granularité trop grossière si elle est retenue provoque inévitablement une perte d'informations pertinentes au moment de l'échantillonnage.

Nous initialisons A_M , la matrice descriptive de \mathcal{M} , comme illustré à la figure 12.

Grâce aux possibilités offertes par l'analyse en composantes principales (ou ACP), nous déduisons *a priori* $\sum_{\bar{x}_M}$, la matrice de variance-covariance, dont nous arrêtons, à l'issue de quelques expériences, le seuil acceptable de convergence vers 0, puis nous identifions les nb indicateurs élémentaires qui sont finalement présents au sein de la chaîne métier observée.

Nous mettons à jour A_M en permutant d'abord l'ensemble des indicateurs de sorte que les nb indicateurs élémentaires se retrouvent en haut à gauche de la matrice descriptive, puis, diagonale mise à part, nous forçons tous les coefficients descriptifs de cette « zone élémentaire » à 0. La figure 13 montre la définition de la « zone élémentaire ».

Grâce à l'ACP toujours, « zone élémentaire » mise à part, nous déduisons *a priori* $\Pi_{\bar{x}_M}$, la matrice de corrélation, puis nous identifions tous les coefficients de corrélation mathématique dits valables, c'est-à-dire dont les valeurs absolues sont supérieures ou égales à $\frac{\sqrt{3}}{2} \approx 0,866$. Dès lors qu'elles tendent vers 0, les indicateurs

5 concernés sont dits indépendants, ce qui permet de compléter la "zone élémentaire".

Nous ordonnons les indicateurs par le biais du calcul *a priori* de $\Gamma_{\bar{x}_M}$, la matrice croisée, qui permet de clarifier les retards croisés des indicateurs corrélés entre eux deux à deux dans la chaîne métier. L'ordonnement s'obtient par le biais de permutations adéquates successives des lignes et des colonnes de la matrice,

10 depuis les indicateurs réalisés en premier.

Nous mettons à jour A_M en forçant à 0 l'ensemble des coefficients descriptifs dépourvus de toute corrélation, puis, en tenant compte des retards croisés, nous permutons les derniers indicateurs non-élémentaires afin d'obtenir une forme générale de la matrice descriptive qui soit la plus proche possible de sa diagonale. La

15 figure 14 montre la forme générale en escalier. La matrice descriptive de M nous renseigne finalement sur la généalogie au sein de la chaîne métier observée. Et ce sont bien les calculs de corrélation qui nous permettent d'approcher au mieux l'ensemble de ces filiations.

Ensuite, de l'interprétation de la matrice descriptive A_M ainsi obtenue, nous

20 déduisons la matrice de connexité A_{nb} . Nous déduisons de la lecture de A_{nb} le nombre de KPI finalement présents dans la chaîne métier observée. Nous identifions clairement les chaînes KPI. Enfin, pour chaque chaîne KPI, nous construisons la matrice descriptive A_N et la matrice de connexité A_m , sans refaire les calculs, en nous appuyant simplement sur les résultats propre à la chaîne métier dans son ensemble.

25 Nous pouvons ensuite passer à l'étape construction du réseau de Petri.

Le calcul des corrélations croisées joue un rôle important dans l'organisation de la matrice descriptive de \mathcal{M} , tout comme celle de \mathcal{M}_{kpi} . Néanmoins, ce calcul n'est pas simple *a priori* et peut même s'avérer impossible dans certain cas, surtout si la chaîne métier est trop complexe. Nous expliquons ci-après la matrice descriptive dès

30 l'obtention des corrélations dites valables, sans qu'il soit nécessaire de passer par les calculs de corrélation croisée, l'ordonnement étant *de facto* exprimé *via* le retour d'expérience, ou aléatoirement le cas échéant. Voici donc les différents cas que nous pouvons rencontrer dans le cas d'une chaîne KPI $\mathcal{M}_{kpi} = \{I_1, \dots, I_m, \dots, I_N\}$:

- le caractère redondant simple sur la figure 15.
- le caractère doublon sur la figure 16.
- le caractère redondant multiple à la figure 17.
- la liaison d'indicateurs à la figure 18.
- 5 - la surcharge d'indicateurs à figure 19.
- la surcharge redondante à figure 20.
- l'interprétation générale à la figure 21.

On automatise ici l'écriture des réseaux de Petri depuis la connaissance des chaînes KPI. Le réseau de Petri est un outil mathématique classique permettant de
10 décrire des relations existant entre des conditions et des événements et de modéliser le comportement de systèmes dynamiques à événements discrets. Un réseau de Petri est en quelque sorte un arbre orienté dit aussi graphe biparti alterné, c'est-à-dire qu'il y a alternance des types de nœuds. Nous nous intéressons ensuite à la circulation des marques au sein de cet arbre orienté. Dans la représentation d'une
15 chaîne KPI, cet arbre orienté prend la forme d'une pyramide.

Considérons deux algorithmes de traitement possibles.

L'algorithme descendant consiste en la construction du réseau de Petri depuis le KPI jusqu'aux indicateurs supposés élémentaires. Cet algorithme sous-entend un
20 « *dés-empilement* » aléatoire et récursif. Lors de chaque cycle, l'algorithme vérifie les connexités entre les indicateurs supposés élémentaires. Mais il faut attendre l'apparition d'indicateurs non corrélés pour que l'algorithme écarte cette proposition du résultat avant de recommencer un nouveau cycle aléatoire et récursif. Cette méthode de traitement est donc peu avantageuse.

L'algorithme montant tient compte, depuis l'ensemble des indicateurs présents
25 au sein de la chaîne KPI, des filiations et donc de la généalogie issue du calcul des corrélations et de l'interprétation de la matrice descriptive. Si l'ordonnancement des indicateurs est maîtrisé, ce qui sous-entend que l'ensemble des corrélations croisées est calculé, et si la complétude des données est vérifiée, ce qui sous-entend que la chaîne KPI est exhaustive au point d'éviter les surcharges d'indicateurs et les caractères redondants multiples, l'algorithme montant autorise alors deux types de
30 réseau de Petri :

- le réseau de Petri parallèle. Soit la chaîne KPI ordonnée $\mathcal{M}_{kpi} = \{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_{10}\}$, et représentée à la figure 22 comme un exemple de réseau de Petri parallèle. On a illustré à la figure 23 la matrice descriptive et les
35 connexités dans le cas du réseau de Petri parallèle.

- le réseau de Petri linéaire. Soit la chaîne KPI ordonnée $\mathcal{M}_{\text{kpi}} = \{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9\}$, et représentée à la figure 24. Il s'agit d'un exemple de réseau de Petri linéaire. La figure 25 montre la matrice descriptive et les connexités dans le cas du réseau de Petri linéaire.

5 Dans ces deux exemples, la matrice de connexité est identique pour un nombre d'indicateurs différent. Dans les deux cas, il s'agit d'un empilement linéaire.

Nous présentons ici le théorème de l'algorithme montant. De façon générale, pour procéder à la génération du réseau de Petri, l'algorithme montant doit opérer par itérations successives $t \leftarrow t+1$, depuis le seuil représenté par l'ensemble des indicateurs élémentaires, sur lequel ont été appliqués les caractères redondants éventuels, en combinant tous les indicateurs élémentaires ou induits d'ores et déjà disponibles, afin de retrouver l'ensemble des indicateurs induits et KPI manquants issus de la superposition de t indicateurs élémentaires. A chaque itération, l'algorithme montant doit s'assurer du bon respect des règles d'intégrité propres aux opérations de redondance ou de superposition. Le processus se poursuit jusqu'à l'obtention définitive du KPI. On note que, à chaque itération t , l'algorithme montant doit préserver l'intégrité des indicateurs déjà obtenus à l'itération $t-1$.

Pour préciser les règles de l'algorithme montant, introduisons maintenant la notion de chaînon, sachant qu'une chaîne KPI est un chaînon complet incluant tous les indicateurs observés dont le KPI. Le principe de l'algorithme montant consiste donc à retrouver, depuis le seuil représenté par l'ensemble des indicateurs élémentaires, l'ensemble des chaînons manquants jusqu'à l'obtention définitive du KPI.

25 Soit un chaînon \mathcal{U} de la chaîne KPI considérée. Ce chaînon nous apporte deux informations :

- un ensemble de données représenté sous la forme d'un triplet ordonné $((...); (...); (...))$ comprenant :

- x indicateurs élémentaires ;

- les liaisons d'indicateurs ;

30 - les caractères redondants.

- une matrice de connexité notée U_x telle que:

- $U_x \in \nabla_{x,x} / \nabla = \{0,1\}$;

- $\forall i, j \in \{1 \dots x\}$, nous associons à toute liaison identifiée et notée L_{ij} le nombre $\delta_{ij} = 1$, ou 0 s'il n'y a pas de liaison sachant que $\delta_{ii} = 0$, $\forall i \in \{1 \dots x\}$;

$$- U_x = [u_{ij}] / u_{ij} = \delta_{ij} = \delta_{ji} = u_{ji} ;$$

- U_x est une matrice carrée creuse symétrique dont la diagonale est toujours nulle.

5 Pour compléter notre démarche de règlementation de l'algorithme montant, nous devons considérer au moins trois chaînons de la chaîne KPI considérée. Aussi, soient deux autres chaînons \mathcal{V} et \mathcal{W} , ayant respectivement pour matrices de connexité : V_y et W_z . Nous considérons maintenant les deux opérations possibles au sein du réseau de Petri.

Soit l'opération de redondance $\mathcal{U} = \mathcal{V} \times \mathcal{C}a_{ij}$:

10 - règle r1 :

\mathcal{V} est induit et intègre (préalablement réalisé), ou élémentaire $\Rightarrow V_y$ est connexe.

- règle r2 :

$$\mathcal{U} = \mathcal{V} \times \mathcal{C}a_{ij} \leftarrow \mathcal{V} \cup (\emptyset; \emptyset; (\mathcal{C}a_{ij})).$$

15 - règle r3 :

\mathcal{U} respecte bien les contraintes issues des calculs de corrélation.

- règle r4 :

U_x est connexe \Rightarrow le graphe orienté de U_x est fortement connexe.

Soit l'opération de superposition $\mathcal{U} = [\mathcal{V} + \mathcal{W}]_{L_{ij}}$ (commutative) :

20 - règle s1 :

- \mathcal{V} est induit et intègre (préalablement réalisé), ou élémentaire $\Rightarrow V_y$ est connexe ;

- \mathcal{W} est induit et intègre (préalablement réalisé), ou élémentaire $\Rightarrow W_z$ est connexe.

25 - règle s2 :

$$\mathcal{V} \cap \mathcal{W} = \emptyset.$$

- règle s3 :

$$\mathcal{U} = [\mathcal{V} + \mathcal{W}]_{L_{ij}} \leftarrow \mathcal{V} \cup \mathcal{W} \cup (\emptyset; (L_{ij}); \emptyset).$$

- règle s4 :

30 \mathcal{U} respecte bien les contraintes issues des calculs de corrélation.

- règle s5 :

U_x est connexe \Rightarrow le graphe orienté de U_x est fortement connexe.

Afin de clarifier les règles de l'algorithme montant, voici une simulation construite sur la base d'une chaîne KPI fictive.

Soit la chaîne KPI ordonnée $\mathcal{M}_{kpi} = \{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_{10}, I_{11}, I_{12}\}$,

5 dont le graphe des liaisons par superposition peut être celui illustré à la figure 26. S'agissant d'une chaîne KPI, l'empilement des indicateurs élémentaires au sein de \mathcal{M}_{kpi} est linéaire comme illustré à la figure 27. La matrice descriptive et les connexités de cette chaîne KPI sont illustrées à la figure 28.

De ces trois représentations, nous pouvons décrire tous les chaînons qui sont nécessaires au bon déroulement de l'algorithme montant lors de la génération du réseau de Petri :

$$- \mathcal{M}_{kpi} = ((I_1, I_2, I_3, I_4, I_5); (L_{12}, L_{23}, L_{34}, L_{45}); (Ca_{r6}, Ca_{r10}, Ca_{r12}));$$

$$- \mathcal{M}_{i11} = ((I_1, I_2, I_3, I_4, I_5); (L_{12}, L_{23}, L_{34}, L_{45}); (Ca_{r6}, Ca_{r10}));$$

$$- \mathcal{M}_{i10} = ((I_1, I_2, I_3); (L_{12}, L_{23}); (Ca_{r10}));$$

$$- \mathcal{M}_{i9} = ((I_1, I_2, I_3); (L_{12}, L_{23}); \emptyset);$$

$$15 \quad - \mathcal{M}_{i8} = ((I_4, I_5); (L_{45}); (Ca_{r6}));$$

$$- \mathcal{M}_{i7} = ((I_1, I_2); (L_{12}); \emptyset);$$

$$- \mathcal{M}_{i6} = ((I_5); \emptyset; (Ca_{r6})).$$

On voit que le principe est le même que celui dit « des poupées russes ». Mais il est préférable d'anticiper l'utilisation informatique en donnant à ces écritures une forme générale qui utilise uniquement les identifiants des indicateurs de la chaîne KPI. Voici donc une écriture plus appropriée à l'exploitation informatique :

$$- \mathcal{M}_{kpi} = ((I_1, I_2, I_3, I_4, I_5); (I_7, I_9, I_{11}, I_8); (I_6, I_{10}, I_{12}));$$

$$- \mathcal{M}_{i11} = ((I_1, I_2, I_3, I_4, I_5); (I_7, I_9, I_{11}, I_8); (I_6, I_{10}));$$

$$- \mathcal{M}_{i10} = ((I_1, I_2, I_3); (I_7, I_9); (I_{10}));$$

$$25 \quad - \mathcal{M}_{i9} = ((I_1, I_2, I_3); (I_7, I_9); \emptyset);$$

$$- \mathcal{M}_{i8} = ((I_4, I_5); (I_8); (I_6));$$

$$- \mathcal{M}_{i7} = ((I_1, I_2); (I_7); \emptyset);$$

$$- \mathcal{M}_{i6} = ((I_5); \emptyset; (I_6)).$$

Cette liste compte en tout 7 chaînons et n'est donc pas suffisante. Nous devons ajouter à cette liste 5 chaînons supplémentaires de type « singleton » puisqu'uniquement constitués des indicateurs élémentaires :

- 5
- $\mathcal{M}_{i_5} = ((I_5); \emptyset; \emptyset)$;
 - $\mathcal{M}_{i_4} = ((I_4); \emptyset; \emptyset)$;
 - $\mathcal{M}_{i_3} = ((I_3); \emptyset; \emptyset)$;
 - $\mathcal{M}_{i_2} = ((I_2); \emptyset; \emptyset)$;
 - $\mathcal{M}_{i_1} = ((I_1); \emptyset; \emptyset)$.

10 Ainsi, voici la liste ordonnée, selon l'algorithme montant, des opérations permettant de construire le réseau de Petri de la chaîne KPI considérée :

- $\mathcal{M}_{i_6} = \mathcal{M}_{i_5} \times I_6$;
- $\mathcal{M}_{i_7} = [\mathcal{M}_{i_1} + \mathcal{M}_{i_2}] I_7$;
- $\mathcal{M}_{i_8} = [\mathcal{M}_{i_4} + \mathcal{M}_{i_6}] I_8$;
- $\mathcal{M}_{i_9} = [\mathcal{M}_{i_3} + \mathcal{M}_{i_7}] I_9$;

15

- $\mathcal{M}_{i_{10}} = \mathcal{M}_{i_9} \times I_{10}$;
- $\mathcal{M}_{i_{11}} = [\mathcal{M}_{i_8} + \mathcal{M}_{i_{10}}] I_{11}$;
- $\mathcal{M}_{i_{12}} = \mathcal{M}_{kpi} = \mathcal{M}_{i_{11}} \times I_{12}$.

Les chaînons des lignes relatives à \mathcal{M}_{i_7} et \mathcal{M}_{i_8} peuvent être réalisés simultanément.

20 Enfin, nous réécrivons maintenant l'ensemble de la chaîne KPI, ce qui revient à reconstituer entièrement « la poupée russe » :

- $\mathcal{M}_{kpi} = \mathcal{M}_{i_{11}} \times I_{12}$;
- $\mathcal{M}_{kpi} = [\mathcal{M}_{i_8} + \mathcal{M}_{i_{10}}] I_{11} \times I_{12}$;
- $\mathcal{M}_{kpi} = [[\mathcal{M}_{i_4} + \mathcal{M}_{i_6}] I_8 + \mathcal{M}_{i_9} \times I_{10}] I_{11} \times I_{12}$;

25

- $\mathcal{M}_{kpi} = [[((I_4); \emptyset; \emptyset) + \mathcal{M}_{i_5} \times I_6] I_8 + [\mathcal{M}_{i_3} + \mathcal{M}_{i_7}] I_9 \times I_{10}] I_{11} \times I_{12}$;
- $\mathcal{M}_{kpi} = [[((I_4); \emptyset; \emptyset) + ((I_5); \emptyset; \emptyset) \times I_6] I_8 \dots$
 $\quad + [((I_3); \emptyset; \emptyset) + [\mathcal{M}_{i_1} + \mathcal{M}_{i_2}] I_7] I_9 \times I_{10}] I_{11} \times I_{12}$;
- $\mathcal{M}_{kpi} = [[((I_4); \emptyset; \emptyset) + ((I_5); \emptyset; \emptyset) \times I_6] I_8 \dots$

$$+ [(I_3); \emptyset; \emptyset) + [(I_1); \emptyset; \emptyset) + ((I_2); \emptyset; \emptyset)]I_7]I_9 \times I_{10}]I_{11} \times I_{12}.$$

Nous pouvons encore simplifier cette écriture en remplaçant les chaînons de type « singleton » par les indicateurs élémentaires. Voici donc l'écriture définitive ou code du réseau de Petri de la chaîne KPI observée :

$$- \mathcal{M}_{kpi} = [[I_4 + I_5 \times I_6]I_8 + [I_3 + [I_1 + I_2]I_7]I_9 \times I_{10}]I_{11} \times I_{12}.$$

Ainsi, voici les trois états successifs par lesquels nous sommes passés lors du traitement des corrélations au sein de la chaîne KPI considérée :

$$- \mathcal{M}_{kpi} = \{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_{10}, I_{11}, I_{12}\};$$

$$- \mathcal{M}_{kpi} = ((I_1, I_2, I_3, I_4, I_5), (I_7, I_9, I_{11}, I_8), (I_6, I_{10}, I_{12}));$$

$$- \mathcal{M}_{kpi} = [[I_4 + I_5 \times I_6]I_8 + [I_3 + [I_1 + I_2]I_7]I_9 \times I_{10}]I_{11} \times I_{12}.$$

Voyons maintenant comment nous pouvons retrouver simplement le réseau de Petri depuis la connaissance de la chaîne KPI.

L'état d'un KPI peut donc être modélisé par une arborescence de type réseau de Petri dont chaque place représente un indicateur dans un état déterminé. En informatique, cette arborescence est codée sous forme d'une chaîne de symboles dite « bien parenthésée », où chaque symbole représente soit un indicateur soit une opération, et chaque niveau de parenthèse dans la chaîne représente une branche particulière du réseau de Petri. Une chaîne « bien parenthésée » est telle que toute parenthèse ouverte est obligatoirement refermée à un autre endroit de la chaîne.

Dans le cas du KPI de l'exemple précédent, le réseau de Petri permettant la modélisation de cet indicateur possède un code unique représenté par la chaîne suivante :

$$\mathcal{M}_{kpi} = [[I_4 + I_5 \times I_6]I_8 + [I_3 + [I_1 + I_2]I_7]I_9 \times I_{10}]I_{11} \times I_{12}.$$

On note que le code du réseau de Petri peut évoluer au cours du temps.

Pour décrire cette chaîne, on utilise une grammaire formelle, aussi connue sous le nom de système de réécriture (une chaîne de caractère qui en remplace une autre). Pour cela, nous devons expliquer la sémantique ainsi trouvée, et nous donner des règles pour décrire toutes les opérations réalisées par les différents symboles de la chaîne lors d'une itération. L'application successive de ces règles à un symbole initial produit une séquence de transformation d'état. Ainsi pour chaque itération, l'état présent du réseau de Petri est représenté par une chaîne de symboles issue de cette séquence.

L'état futur au pas d'itération suivant est calculé en transformant simultanément tous les symboles de l'état présent à l'aide des règles spécifiées par la grammaire suivante :

- définition des symboles propres à la chaîne réseau de Petri :

- 5
- l'espace est muet ;
 - I_i désigne un indicateur de la chaîne KPI ;
 - + désigne une opération de superposition ;
 - x désigne une opération de redondance ;
 - [désigne l'ouverture d'une superposition ;
 - 10
 -] désigne la fermeture d'une superposition ;
 - U désigne une opération entre deux chaînons ;
 - ((...);(...);(...)) désigne un chaînon ;
 - \emptyset désigne l'absence d'information dans cette partie du chaînon.

- définition des règles ordonnées permettant l'évolution au pas d'itération suivant :

15

- règle (a) : les indicateurs élémentaires présents dans la chaîne réseau de Petri sont remplacés par les chaînons de type « *singleton* » ; les indicateurs élémentaires sont facilement identifiables puisque toujours placés après les symboles [et +. Cette règle ne s'applique qu'une fois en début de traitement ;

20

$$I_i \Rightarrow ((I_i); \emptyset; \emptyset)$$

- règle (b) : le seuil de traitement de l'algorithme montant est mis à jour avant chaque itération en remplaçant tous les caractères redondants appliqués uniquement sur les indicateurs élémentaires ou intègres, par les opérations de redondance correspondantes. Les caractères redondants concernés sont représentés par x I_i et sont toujours placés après le symbole), la règle (a) ayant été réalisée préalablement à la règle (b) en début de traitement ;

25

$$x I_i \Rightarrow U (\emptyset; \emptyset; (I_i))$$

- règle (c) : à chaque itération depuis le seuil de traitement, nous réalisons toutes les liaisons d'indicateurs les plus encapsulées dans la chaîne réseau de Petri en les remplaçant par les opérations de superposition correspondantes. Les liaisons d'indicateurs concernées sont représentées par l'enchaînement « *bien parenthésé* » [...+...] I_i , la règle (b) ayant toujours été réalisée préalablement à la règle (c) ;

30

$$[...+...] I_i \Rightarrow \dots U \dots U (\emptyset; (I_i); \emptyset)$$

- règle (d) : c'est la règle des « *poupées russes* » dite aussi de simplification de l'écriture des chaînons. A la fin de chaque itération, les opérations de redondance

35

ou de superposition sont remplacées par les chaînons résultats. Les opérations concernées sont représentées par les enchainements de type $((I);(J);(K)) \cup ((L);(M);(N))$ pour la redondance ou $((I);(J);(K)) \cup ((L);(M);(N)) \cup ((O);(P);(Q))$ pour la superposition ;

$$5 \quad ((I);(J);(K)) \cup ((L);(M);(N)) \Rightarrow ((I,L);(J,M);(K,N))$$

Appliquons ces règles de grammaire des graphes sur la chaîne réseau de Petri obtenue lors de la simulation précédente :

$$\mathcal{M}_{kpi} = [I_4 + I_5 \times I_6] I_8 + [I_3 + [I_1 + I_2] I_7] I_9 \times I_{10}] I_{11} \times I_{12}$$

Les réécritures successives sont donc les suivantes:

10 Selon la règle (a):

$$[[((I_4);\emptyset;\emptyset) + ((I_5);\emptyset;\emptyset) \times I_6] I_8 + [((I_3);\emptyset;\emptyset) + [((I_1);\emptyset;\emptyset) + ((I_2);\emptyset;\emptyset)] I_7] I_9 \times I_{10}] I_{11} \times I_{12}$$

Selon les règles (b)+(d):

$$15 \quad [[((I_4);\emptyset;\emptyset) + ((I_5);\emptyset;(I_6))] I_8 + [((I_3);\emptyset;\emptyset) + [((I_1);\emptyset;\emptyset) + ((I_2);\emptyset;\emptyset)] I_7] I_9 \times I_{10}] I_{11} \times I_{12}$$

Selon les règles (c)+(d):

$$[((I_4, I_5);(I_8);(I_6)) + [((I_3);\emptyset;\emptyset) + ((I_1, I_2);(I_7);\emptyset)] I_9 \times I_{10}] I_{11} \times I_{12}$$

Selon les règles (c)+(d):

$$[((I_4, I_5);(I_8);(I_6)) + ((I_1, I_2, I_3);(I_7, I_9);\emptyset) \times I_{10}] I_{11} \times I_{12}$$

20 Selon les règles (b)+(d):

$$[((I_4, I_5);(I_8);(I_6)) + ((I_1, I_2, I_3);(I_7, I_9);(I_{10}))] I_{11} \times I_{12}$$

Selon les règles (c)+(d):

$$((I_1, I_2, I_3, I_4, I_5);(I_7, I_9, I_{11}, I_8);(I_6, I_{10})) \times I_{12}$$

Selon les règles (b)+(d):

$$25 \quad ((I_1, I_2, I_3, I_4, I_5);(I_7, I_9, I_{11}, I_8);(I_6, I_{10}, I_{12}))$$

Le réseau de Petri correspondant à ce résultat est illustré à la figure 29.

Le fait de pouvoir représenter les chaînes KPI sous forme de réseau de Petri présente donc deux gros avantages. Il permet le renforcement de la qualité des données au sein de l'entrepôt de données. Et il permet l'analyse quantitative des données grâce aux évolutions techniques qu'autorise ce type de représentation que ce soit en termes de réseau de Petri dits temporisés, interprétés, stochastiques, colorés, continus et/ou hybrides.

30

Ici la compréhension du modèle de circulation des marques au sein du réseau de Petri retient notre attention. Ces marques sont la représentation des réalisations

ordonnées au sein du système observé. Les valeurs transportées par ces marques autorisent une lecture des distributions au sein du système d'information.

En outre, une telle représentation, combinée à des outils de type « *files d'attente* » trouve alors tout son sens.

5 Au sein d'une organisation telle qu'une société commerciale, le procédé qui vient d'être présenté peut être mis en œuvre avec les moyens illustrés à la figure 30.

10 Un dispositif 2 tel qu'un serveur est relié à au moins une base de données 4 d'un entrepôt de données (la base pouvant former ce dernier). La base contient les indicateurs relatifs à au moins un objet tel que l'organisation elle-même ou des activités de celle-ci, cet objet prenant plusieurs états (par exemple à des date successives), à chaque indicateur correspondant des valeurs associées aux états respectifs.

15 Le dispositif 2 comprend notamment une unité centrale et une mémoire formant des moyens aptes à identifier, comme expliqué plus haut, des corrélations entre au moins certains des indicateurs. On considérera ainsi par exemple trois indicateurs I_1 , I_2 et I_3 .

20 Les moyens sont aptes à former des groupes tels que des chaînes constitués chacun par un indicateur et des indicateurs corrélés à ce dernier et suffisant pour déduire ce dernier. Ici, on suppose que I_1 et I_2 induisent I_3 . Il s'agit donc du groupe (I_1, I_2, I_3) .

Les moyens sont aptes à construire le réseau de Petri du groupe et à le coder au moyen de la grammaire comme indiqué plus haut.

25 Ces moyens sont aptes à associer, dans la base 4, à chacun des indicateurs I_1 , I_2 et I_3 du groupe, dans un champs correspondant, une donnée M qui est la codification obtenue et qui indique une identité des autres indicateurs du groupe et leurs relations de cause à effet. On a aussi porté sur la figure 30 une autre donnée N relative à un autre groupe concernant I_1 et I_2 mais pas I_3 .

30 C'est ainsi que l'on constitue la base de données 4 puis qu'on la maintient à jour. En effet, lorsque de nouveaux indicateurs sont créés, des corrélations sont recherchées et la mise en œuvre du procédé enrichit la base.

L'exploitation de la base peut se faire comme suit.

35 On suppose qu'un opérateur tel qu'une personne consulte le dispositif 2 depuis son ordinateur 6 relié à un réseau de télécommunication 8 tel qu'internet ou un intranet, via un serveur 10 de l'organisation. Il souhaite par exemple connaître une valeur récente de l'indicateur I_1 .

Le dispositif 2 identifie dans la base 4 les données M et N indiquant une identité des indicateurs I_2 et I_3 corrélés à l'indicateur consulté I_1 . Le dispositif déduit de la codification au moyen de la grammaire les relations de cause à effet entre les indicateurs du groupe M. Il fait de même avec ceux du groupe N. Le dispositif est aussi apte à exploiter ces identités, par exemple en transmettant ces identités au dispositif 10 chargé de construire et de modifier une page web 12 consultée par l'opérateur, afin que le dispositif 10 y fasse apparaître les identités des indicateurs identifiés et leurs relations.

L'opérateur est ainsi informé, après avoir consulté l'indicateur I_1 sur la page, qu'il devrait aussi être attentif à la valeur des indicateurs I_2 et I_3 qui lui sont corrélés.

On met donc en œuvre un procédé dans lequel:

- les moyens 2 identifient des corrélations entre au moins certains des indicateurs;

- les moyens 2 forment des groupes constitués chacun par un indicateur et des indicateurs corrélés à ce dernier et suffisant pour déduire ce dernier;

- les moyens associent, dans la base 4, à chacun des indicateurs du groupe, une donnée indiquant une identité des autres indicateurs du groupe.

Dans ce même procédé :

- les moyens identifient dans la base 4 au moins une donnée M et N indiquant une identité des indicateurs I_2 et I_3 corrélés à un indicateur consulté I_1 et formant avec lui un groupe tel que les indicateurs du groupe, sauf un, suffisent pour déduire ce dernier, et

- les moyens exploitent les identités des indicateurs identifiés, par exemple en les transmettant sur le réseau 8.

Les dispositifs 2 et 10 comprennent un ou plusieurs programmes d'ordinateur comprenant des instructions de code apte à commander la mise en œuvre de certaines au moins des étapes du procédé lorsqu'il est exécuté sur un ordinateur.

On peut mettre ce programme sur un support d'enregistrement tel qu'un disque DVD, un disque dur ou une mémoire flash. On peut mettre ce programme à disposition sur un réseau de télécommunication tel que le réseau 8 en vue de son téléchargement.

L'invention permet d'aller encore plus loin en réalisant une surveillance ou monitoring en trois dimensions des variables, comme nous allons maintenant le montrer.

On vient de voir que la CsKPI, ou codification significative des KPI, permet la construction de deux outils d'analyse et d'aide à la décision, à savoir la matrice

descriptive du système complexe observé et le makespan ou réseau de Petri temporisé, formant un indicateur composite de robustesse.

Ces deux représentations permettent de produire l'attribut de la qualité des données de supervision, qui reste sans doute l'attribut le plus difficile à conceptualiser, à savoir l'intégrité, c'est-à-dire aussi l'authenticité et la traçabilité. La CsKPI permet la supervision des points de fragilité du système complexe observé. Elle permet ainsi de localiser précisément des incidents, d'en suivre les propagations, les ricochets, mais aussi d'en mesurer avec précision les temps de propagation. Elle autorise enfin la simulation, la stimulation et l'optimisation du modèle de supervision.

Mais la CsKPI permet aussi de produire un autre outil d'analyse et d'aide à la décision qui est une surveillance en trois dimensions des variables. Cette surveillance permet d'évaluer la complétude d'un outil de supervision vis-à-vis du système complexe observé. La complétude est également l'un des attributs de la qualité des données de supervision, tels que sont listés plus haut.

Nous allons donc décrire d'autres étapes du procédé dans lesquelles :

- on sélectionne l'une I_0 des variables et on place un vecteur correspondant à la variable;
 - on sélectionne l'une I_1 des variables qui lui est le plus étroitement corrélée et on place un vecteur correspondant à cette variable à une position représentant cette corrélation ;
 - puis, autant de fois que nécessaire, on sélectionne l'une I_i des variables qui est le plus étroitement corrélée à celles correspondant à un vecteur déjà positionné et on place un vecteur correspondant à cette variable à une position représentant cette corrélation,
- les vecteurs étant positionnés avec leur origine au centre d'une sphère représentant l'objet et leur extrémité sur la sphère.

En effet, dans l'outil de surveillance des variables, le système complexe observé prend la forme d'une sphère et le système de supervision prend la forme d'un maillage, par exemple de type géodésique, qui s'enroule progressivement autour de la sphère.

On rappelle qu'il est possible de représenter les variables présentes au sein de l'entrepôt de données de l'entreprise sous forme de vecteurs d'observations variables au cours du temps. Ces vecteurs sont positionnés les uns par rapport aux autres à l'aide des coefficients de corrélation mathématique qui sont ici les cosinus directeurs des vecteurs d'observations construits deux à deux.

En réalité, cette construction devrait être hypersphérique compte tenu du fait que nous travaillons dans un espace euclidien de dimension équivalente au cardinal de l'entrepôt de données, à savoir le nombre de variables de supervision. Pour des raisons techniques, nous choisissons en l'espèce d'aplatir cette hypersphère afin de travailler uniquement dans un espace à trois dimensions.

Nous allons expliquer le calcul des coordonnées des vecteurs et rappeler tout d'abord les relations de trigonométrie « sphérique ».

Soit un repère orthonormé direct Oxyz. Un triangle sphérique est la plus petite portion d'une sphère limitée par trois grands cercles centrés en O, le centre de la sphère, et ayant même rayon que celui de la sphère. Les côtés de ce triangle sont des courbes dont les prolongements sont des cercles de même rayon que celui de la sphère, comme illustré à la figure 31. Les trois points A, B, C sur la sphère délimitent le triangle « sphérique ».

La figure illustre la sphère dans sa configuration initiale : B est au pôle situé sur l'axe z, A est l'un des points d'intersection de la sphère avec le plan passant par les axes z et x.

Les sommets du triangle sont notés A, B et C. Ainsi l'angle ABC est noté B, CAB est noté A, etc.. Le coté opposé à un angle a la même dénomination que celui-ci mais en minuscule. Le côté BC est ainsi nommé a.

S'agissant d'une sphère de rayon 1, nous pouvons confondre un côté et un angle, ainsi BC = a = rayon BOC = BOC (abscisse curviligne) dès lors que BOC est inférieur à π/6. Cela peut se démontrer facilement par le biais du développement limité de Taylor pour la fonction sinus à l'ordre 2 :

$$abscisse_curviligne = r.\sin x = r.(x + O_{\sin}(x^2))$$

$$\text{où } O_{\sin}(x^2) = -\frac{x^3}{3!} + \sum_{n=2}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

$$\text{d'où } O_{\sin}(x^2) \approx 0 \quad \forall x \in [0, \frac{\pi}{6}]$$

Mais l'angle A est différent de a et ainsi de suite.

Pour faciliter les calculs, nous parlons donc en radians (π rad = 180°). Notre but est de calculer les coordonnées x, y, z de C, un point quelconque de la sphère. Nous les calculons ensuite depuis un repère différent. Ces relations doivent être celles de la trigonométrie « sphérique ».

Dans la suite, on utilise deux figures correspondant à deux projections. Cela nous permet d'avoir les coordonnées en z et en y sur la figure 32, et en x sur la figure

33. Ces projections ne changent pas les valeurs des angles A, B, C, a, b et c, ou tout du moins de manière très négligeable tant que les valeurs de a, b et c restent inférieures à $\pi/6$ et que le point B reste au pôle.

On fait remarquer au passage que $\cos(\pi/6)$ vaut à peu près 0,866 et correspond aussi au seuil des corrélations mathématiques dites remarquables au sens de la matrice descriptive.

A étant un point sur la sphère et dans le plan zOx, la projection de C sur le plan xOy est nommée C'. Pour trouver les coordonnées x et y, nous devons d'abord trouver la valeur de OC'. Cela revient à réaliser un calcul de trigonométrie normale depuis les schémas des figures 2 et 3.

La figure 32 nous montre un simple triangle à l'issue de la projection. Toutefois, la projection de la figure 2 n'est pas tout à fait exacte. En fait, ce n'est pas le plan yOz, mais le plan COB, ce qui n'influe en rien sur les calculs compte tenu des hypothèses sur les angles.

Les formules de z et OC' sont donc:

$$z = \sin(C'OC) = \sin(\pi/2 - a) = \cos a$$

$$OC' = \cos(C'OC) = \cos(\pi/2 - a) = \sin a$$

Pour les coordonnées de C, nous devons calculer les coordonnées de C' dans le plan xOy. Sachant que la longueur de l'hypoténuse est sin a, nous en déduisons les coordonnées du point C :

$$x = OC' \cos(\angle ABC) = \sin a \cos B$$

$$y = OC' \sin(\angle ABC) = \sin a \sin B \quad (1)$$

$$z = \cos a$$

Nous réalisons ensuite un changement de base de sorte que ce soit A qui devienne le pôle par une rotation d'axe Oy, comme illustré à la figure 34.

Nous avons un nouveau repère orthonormé direct Ox'y'z'. Les figures 35 et 36 montrent les coordonnées y et z de C dans le nouveau repère, qui s'expriment comme suit:

$$z' = \sin(C'OC) = \sin(\pi/2 - b) = \cos b$$

$$OC' = \cos(C'OC) = \cos(\pi/2 - b) = \sin b$$

Le point A à la figure 35 n'étant pas du même côté que le point B à la figure 36, nous devons utiliser les cos ou sin ($\pi - A$). D'où les relations suivantes :

$$x' = OC' \cos(\angle XAC) = \sin b \cos(\pi - A) = -\sin b \cos A$$

$$y' = OC' \sin(\angle XAC) = \sin b \sin(\pi - A) = \sin b \sin A \quad (2)$$

$$z' = \cos b$$

Pour établir les relations de trigonométrie sphériques, nous utilisons les relations de changement de repère suivantes :

$$x = x' \cos c + z' \sin c$$

$$y = y'$$

5

$$z = z' \cos c - x' \sin c$$

A partir de (1) et (2) et en remplaçant x' , y' et z' dans les formules de changement de base, on obtient :

$$x = \cos b \sin c - \sin b \cos c \cos A$$

$$y = \sin b \sin A$$

10

$$z = \cos b \cos c + \sin b \sin c \cos A$$

Les formules de trigonométrie « sphérique » sont obtenues en remplaçant maintenant x , y et z :

$$\sin a \cos B = \cos b \sin c - \sin b \cos c \cos A$$

$$\sin a \sin B = \sin b \sin A \tag{3}$$

15

$$\cos a = \cos b \cos c + \sin b \sin c \cos A$$

Pour le calcul des coordonnées des vecteurs, il s'agit d'abord de localiser les variables corrélées à une variable de référence notée I_j et qui a été préalablement sélectionnée par l'utilisateur. L'ordonnancement des variables doit tenir compte de la CsKPI correspondante.

20

Les vecteurs utilisés ont tous pour point d'origine le centre de la sphère en $(0,0,0)_0$ dans la base que nous notons $\{O, X_0, Y_0, Z_0\}$, alors que leurs extrémités sont toutes situées à la surface de la sphère préalablement tracée.

Nous appliquons une méthode de triangulation. Nous plaçons un premier vecteur noté \vec{I}_0 , puis le vecteur noté \vec{I}_1 correspondant à la variable qui lui est la plus fortement corrélée positivement au sens de la CsKPI. Puis nous déduisons \vec{I}_2 par le biais du calcul. Ce troisième vecteur est lui aussi choisi par rapport à son fort niveau de corrélation au sens de la CsKPI et il doit être placé en fonction des deux autres vecteurs déjà positionnés selon une rotation antihoraire autour de \vec{I}_0 .

25

S'ensuit un changement de repère. En effet, la modification du pôle de \vec{I}_i vers $\vec{I}_{(i+1)}$ permet de répéter la méthode de triangulation autant de fois que nécessaire tout en respectant les hypothèses de la trigonométrie « sphérique », à savoir des angles toujours compris entre 0 et $\pi/6$.

30

Avec ce système, les points correspondant aux variables de supervision se dessinent progressivement autour de la sphère à la façon d'une pelure de pomme comme illustré à la figure 37.

Le premier vecteur \vec{I}_0 est positionné sur l'axe des Z du repère orthonormé de la figure 38. La sphère possédant un rayon de 1, le premier vecteur a une norme de 1 avec les coordonnées cartésiennes (0,0,1). Ce vecteur constitue donc le premier pôle de notre construction. Le vecteur \vec{I}_1 représente la variable la plus fortement corrélée positivement au sens de la CsKPI par rapport à la variable de référence, et pour simplifier le calcul de ses coordonnées, nous plaçons son extrémité sur le plan xOz.

Nous pouvons alors poser le problème depuis le cercle trigonométrique de la figure 39. α_{01} est l'angle séparant les deux vecteurs et le cosinus de cet angle est en réalité le cosinus directeur ou le coefficient de corrélation contenu dans la matrice des corrélations.

Nous en déduisons les coordonnées cartésiennes du vecteur \vec{I}_1 :

$$\left(\vec{I}_1\right)_0 = \begin{pmatrix} \sqrt{1-r_{p01}^2} \\ 0 \\ r_{p01} \end{pmatrix}_{X_0Y_0Z_0} = \begin{pmatrix} \sqrt{1-r_{p01}^2} \\ 0 \\ r_{p01} \end{pmatrix}_0$$

r_{p01} est le cosinus directeur contenu dans la matrice des corrélations, positif et compris dans l'intervalle [0.866 , 1]. Ce doit être une corrélation remarquable au sens de la matrice descriptive.

Le troisième vecteur \vec{I}_2 représente la variable la plus fortement corrélée positivement au sens de la CsKPI, et ce respectivement par rapport à \vec{I}_0 et \vec{I}_1 . C'est ici qu'intervient la méthode de triangulation. Le placement de \vec{I}_2 est fait à partir des positions de \vec{I}_0 et \vec{I}_1 comme illustré à la figure 40.

Pour le calcul de \vec{I}_2 , les relations de trigonométrie « sphérique » précitées sont utilisées :

$$\left(\vec{I}_2\right)_0 = \begin{pmatrix} x_{02} \\ y_{02} \\ z_{02} \end{pmatrix}_{X_0Y_0Z_0} = \begin{pmatrix} \sin a \cos B \\ \sin a \sin B \\ \cos a \end{pmatrix}_0$$

Compte-tenu des formules (3), on obtient:

$$\cos A = \frac{\cos a - \cos b \cos c}{\sin b \sin c}$$

$$\sin A = \sqrt{1 - \cos^2 A}$$

$$\sin A = \sqrt{1 - \left(\frac{\cos a - \cos b \cos c}{\sqrt{1 - \cos^2 b} \sqrt{1 - \cos^2 c}} \right)^2}$$

$$\sin B = \frac{\sin b \sin A}{\sin a}$$

5

$$\left\{ \begin{aligned} \sin B &= \frac{\sqrt{1 - \cos^2 b}}{\sqrt{1 - \cos^2 a}} \sqrt{1 - \left(\frac{\cos a - \cos b \cos c}{\sqrt{1 - \cos^2 b} \sqrt{1 - \cos^2 c}} \right)^2} \\ \cos B &= \sqrt{1 - \sin^2 B} \end{aligned} \right.$$

$$\begin{pmatrix} \vec{I}_2 \end{pmatrix}_0 = \begin{pmatrix} x_{02} \\ y_{02} \\ z_{02} \end{pmatrix}_{X_0 Y_0 Z_0} = \begin{pmatrix} \sqrt{1 - \cos^2 a} \cos B \\ \sqrt{1 - \cos^2 a} \sin B \\ \cos a \end{pmatrix}_0$$

10

$$\begin{pmatrix} \vec{I}_2 \end{pmatrix}_0 = \begin{pmatrix} \sqrt{1 - \cos^2 a} \sqrt{1 - \sin^2 B} \\ \sqrt{1 - \cos^2 b} \sqrt{1 - \left(\frac{\cos a - \cos b \cos c}{\sqrt{1 - \cos^2 b} \sqrt{1 - \cos^2 c}} \right)^2} \\ \cos a \end{pmatrix}_0$$

$$\begin{pmatrix} \vec{I}_2 \end{pmatrix}_0 = \begin{pmatrix} \sqrt{1 - \cos^2 a} \sqrt{1 - \frac{1 - \cos^2 b}{1 - \cos^2 a} \left(1 - \left(\frac{\cos a - \cos b \cos c}{\sqrt{1 - \cos^2 b} \sqrt{1 - \cos^2 c}} \right)^2 \right)} \\ \sqrt{(1 - \cos^2 b) - \left(\frac{\cos a - \cos b \cos c}{\sqrt{1 - \cos^2 c}} \right)^2} \\ \cos a \end{pmatrix}_0$$

$$\begin{pmatrix} \sqrt{(1 - \cos^2 a) - \left((1 - \cos^2 b) - \frac{(\cos a - \cos b \cos c)^2}{1 - \cos^2 c} \right)} \\ \sqrt{1 - \cos^2 b - \frac{(\cos a - \cos b \cos c)^2}{1 - \cos^2 c}} \\ \cos a \end{pmatrix}_0$$

$$\begin{pmatrix} \sqrt{\cos^2 b - \cos^2 a + \frac{(\cos a - \cos b \cos c)^2}{1 - \cos^2 c}} \\ \sqrt{1 - \cos^2 b - \frac{(\cos a - \cos b \cos c)^2}{1 - \cos^2 c}} \\ \cos a \end{pmatrix}_0$$

Il ne reste plus qu'à utiliser l'ensemble des cosinus directeurs liés à α_{01} , α_{02} α_{12} depuis la matrice des corrélations, comme illustré à la figure 41.

5 Le cosinus de l'angle α_{01} qui sépare les vecteurs \vec{I}_0 et \vec{I}_1 est donc situé dans la matrice des corrélations dans la case correspondant aux deux variables associées. Il doit être positif et compris dans l'intervalle [0.866 , 1]. Ce doit être une corrélation remarquable au sens de la matrice descriptive.

10 Nous notons cette valeur r_{p01} , sachant que par définition, $r_{p01} = r_{p10}$. Nous faisons de même pour chacun des angles directeurs concernés.

Cela nous permet d'écrire les coordonnées cartésiennes du vecteur \vec{I}_2 , exprimées dans la base d'origine {O, X₀, Y₀, Z₀}, de la façon suivante :

$$\begin{pmatrix} \sqrt{r_{p12}^2 - r_{p02}^2 + \frac{(r_{p02} - r_{p01} r_{p12})^2}{1 - r_{p01}^2}} \\ \sqrt{1 - r_{p12}^2 - \frac{(r_{p02} - r_{p01} r_{p12})^2}{1 - r_{p01}^2}} \\ r_{p02} \end{pmatrix}_0$$

Avant de positionner les vecteurs suivants, un changement de base doit être effectué. En effet, jusqu'à présent, nous avons pour pôle le vecteur origine \vec{I}_0 . Nous devons désormais basculer le pôle sur le vecteur \vec{I}_1 . Une modification du repère apparaît et il faut alors utiliser les matrices de passage successives pour continuer à exprimer les coordonnées des vecteurs par rapport à la base d'origine {O,X₀,Y₀,Z₀}.

La nouvelle base se note alors {O,X₁,Y₁,Z₁} et la méthode de triangulation s'opère désormais sur les variables I_1 , I_2 et I_3 , comme illustré à la figure 42.

Plus généralement, nous effectuons un changement de pôle de \vec{I}_i vers $\vec{I}_{(i+1)}$ pour placer $\vec{I}_{(i+3)}$ et ce dernier est alors exprimé dans un premier temps dans la nouvelle base {O, X_(i+1), Y_(i+1), Z_(i+1)} avant d'être exprimé dans la base {O, X₀, Y₀, Z₀}.

Après calcul, nous pouvons écrire les coordonnées de \vec{I}_3 qui sont exprimées dans la nouvelle base {O, X₁, Y₁, Z₁} : $(\vec{I}_3)_1$. Nous effectuons alors le même calcul que pour le placement de \vec{I}_2 et nous trouvons les coordonnées cartésiennes du vecteur \vec{I}_3 simplement en remplaçant les indices :

$$(\vec{I}_3)_1 = \begin{pmatrix} \sqrt{r_{p23}^2 - r_{p13}^2 + \frac{(r_{p13} - r_{p12} r_{p23})^2}{1 - r_{p12}^2}} \\ \sqrt{1 - r_{p23}^2 - \frac{(r_{p13} - r_{p12} r_{p23})^2}{1 - r_{p12}^2}} \\ r_{p13} \end{pmatrix}_1$$

Pour exprimer le vecteur \vec{I}_3 dans notre base d'origine {O, X₀, Y₀, Z₀}, nous utilisons la matrice de passage P₁ qui est constituée des vecteurs unitaires de {O, X₁, Y₁, Z₁} exprimés dans la base d'origine {O, X₀, Y₀, Z₀}.

Nous exprimons alors $(\vec{I}_3)_0$ dans la base d'origine de la façon suivante :

$$(\vec{I}_3)_0 = \begin{pmatrix} x_{03} \\ y_{03} \\ z_{03} \end{pmatrix}_{X_0 Y_0 Z_0} = P_1 [01] \begin{pmatrix} x_{13} \\ y_{13} \\ z_{13} \end{pmatrix}_{X_1 Y_1 Z_1} = P_1 [01] (\vec{I}_3)_1$$

Pour le positionnement du vecteur \vec{I}_i , après une recherche de récurrence depuis les formules de positionnement des vecteurs, nous déduisons que, pour tout $i \geq 3$, les coordonnées cartésiennes du vecteur \vec{I}_i sont :

$$\begin{aligned}
 (\vec{I}_i)_{(i-2)} &= \begin{pmatrix} x_{(i-2)i} \\ y_{(i-2)i} \\ z_{(i-2)i} \end{pmatrix}_{X_{(i-2)}Y_{(i-2)}Z_{(i-2)}} \\
 5 \quad (\vec{I}_i)_{(i-2)} &= \begin{pmatrix} \sqrt{r_{p(i-1)i}^2 - r_{p(i-2)i}^2 + \frac{(r_{p(i-2)i} - r_{p(i-2)(i-1)}r_{p(i-1)i})^2}{1 - r_{p(i-2)(i-1)}^2}} \\ \sqrt{1 - r_{p(i-1)i}^2 - \frac{(r_{p(i-2)i} - r_{p(i-2)(i-1)}r_{p(i-1)i})^2}{1 - r_{p(i-2)(i-1)}^2}} \\ r_{p(i-2)i} \end{pmatrix}_{X_{(i-2)}Y_{(i-2)}Z_{(i-2)}}
 \end{aligned}$$

Il est entendu que ces coordonnées ne sont pas encore exprimées dans la base d'origine.

Il est indispensable d'utiliser des matrices de passage successives pour retrouver les coordonnées de nos vecteurs successifs \vec{I}_i dans la base d'origine $\{O, X_0, Y_0, Z_0\}$. Il s'agit donc de déterminer les coordonnées de la nouvelle base $\{O, X_{(i-2)}, Y_{(i-2)}, Z_{(i-2)}\}$ dans la base d'origine $\{O, X_0, Y_0, Z_0\}$. Pour ce faire, on procède par récurrence. P_1 étant la première matrice de passage rencontrée, nous devons l'utiliser comme amorce de notre récurrence pour le cas où $i = 3$. P_1 correspond donc à la matrice de passage depuis la base $\{O, X_0, Y_0, Z_0\}$ vers la base $\{O, X_1, Y_1, Z_1\}$.

Voici le premier vecteur de cette base que nous pouvons déduire facilement :

$$z_1 = (\vec{I}_1)_0 = \begin{pmatrix} x_{01} \\ y_{01} \\ z_{01} \end{pmatrix}_0$$

Le deuxième vecteur de cette base se déduit du produit vectoriel suivant :

$$y_1 = \frac{1}{\sin \alpha_{12}} (\vec{I}_1)_0 \wedge (\vec{I}_2)_0 = \frac{1}{\sin \alpha_{12}} \begin{pmatrix} x_{01} \\ y_{01} \\ z_{01} \end{pmatrix}_0 \wedge \begin{pmatrix} x_{02} \\ y_{02} \\ z_{02} \end{pmatrix}_0 = \frac{1}{\sqrt{1-r_{p12}^2}} \begin{pmatrix} y_{01}z_{02} - z_{01}y_{02} \\ z_{01}x_{02} - x_{01}z_{02} \\ x_{01}y_{02} - y_{01}x_{02} \end{pmatrix}_0$$

Le troisième vecteur de cette base se déduit du produit vectoriel suivant :

$$x_1 = y_1 \wedge z_1 = \frac{1}{\sqrt{1-r_{p12}^2}} \begin{pmatrix} y_{01}z_{02} - z_{01}y_{02} \\ z_{01}x_{02} - x_{01}z_{02} \\ x_{01}y_{02} - y_{01}x_{02} \end{pmatrix}_0 \wedge \begin{pmatrix} x_{01} \\ y_{01} \\ z_{01} \end{pmatrix}_0$$

$$5 \quad x_1 = \frac{1}{\sqrt{1-r_{p12}^2}} \begin{pmatrix} z_{01}(z_{01}x_{02} - x_{01}z_{02}) - y_{01}(x_{01}y_{02} - y_{01}x_{02}) \\ x_{01}(x_{01}y_{02} - y_{01}x_{02}) - z_{01}(y_{01}z_{02} - z_{01}y_{02}) \\ y_{01}(y_{01}z_{02} - z_{01}y_{02}) - x_{01}(z_{01}x_{02} - x_{01}z_{02}) \end{pmatrix}_0$$

Nous en déduisons la matrice de passage correspondante :

$$P_1[01] = (x_1 \quad y_1 \quad z_1)_{01}$$

Nous pouvons ainsi déduire facilement la formule de récurrence pour le cas

$i \geq 3$ et \vec{I}_i exprimé dans la base d'origine:

$$10 \quad (\vec{I}_i)_0 = \begin{pmatrix} x_{0i} \\ y_{0i} \\ z_{0i} \end{pmatrix}_{X_0Y_0Z_0} = P_{(i-2)}[0(i-2)] \begin{pmatrix} x_{(i-2)i} \\ y_{(i-2)i} \\ z_{(i-2)i} \end{pmatrix}_{X_{(i-2)}Y_{(i-2)}Z_{(i-2)}} = P_{(i-2)}[0(i-2)] (\vec{I}_i)_{(i-2)}$$

$$\text{où } P_{(i-2)}[0(i-2)] = (x_{(i-2)} \quad y_{(i-2)} \quad z_{(i-2)})_{0(i-2)}$$

$$z_{(i-2)} = (\vec{I}_{(i-2)})_0 = \begin{pmatrix} x_{0(i-2)} \\ y_{0(i-2)} \\ z_{0(i-2)} \end{pmatrix}_0$$

$$y_{(i-2)} = \frac{1}{\sin \alpha_{(i-2)(i-1)}} (\vec{I}_{(i-2)})_0 \wedge (\vec{I}_{(i-1)})_0 = \frac{1}{\sqrt{1-r_{p(i-2)(i-1)}^2}} \begin{pmatrix} y_{0(i-2)}z_{0(i-1)} - z_{0(i-2)}y_{0(i-1)} \\ z_{0(i-2)}x_{0(i-1)} - x_{0(i-2)}z_{0(i-1)} \\ x_{0(i-2)}y_{0(i-1)} - y_{0(i-2)}x_{0(i-1)} \end{pmatrix}_0$$

$$x_{(i-2)} = y_{(i-2)} \wedge z_{(i-2)}$$

$$15 \quad x_{(i-2)} = \frac{1}{\sqrt{1-r_{p(i-2)(i-1)}^2}} \begin{pmatrix} z_{0(i-2)}(z_{0(i-2)}x_{0(i-1)} - x_{0(i-2)}z_{0(i-1)}) - y_{0(i-2)}(x_{0(i-2)}y_{0(i-1)} - y_{0(i-2)}x_{0(i-1)}) \\ x_{0(i-2)}(x_{0(i-2)}y_{0(i-1)} - y_{0(i-2)}x_{0(i-1)}) - z_{0(i-2)}(y_{0(i-2)}z_{0(i-1)} - z_{0(i-2)}y_{0(i-1)}) \\ y_{0(i-2)}(y_{0(i-2)}z_{0(i-1)} - z_{0(i-2)}y_{0(i-1)}) - x_{0(i-2)}(z_{0(i-2)}x_{0(i-1)} - x_{0(i-2)}z_{0(i-1)}) \end{pmatrix}_0$$

Cette formulation produit donc les règles de construction de l'outil complémentaire permettant la surveillance en trois dimensions des variables.

Comme illustré aux figures 43 et 44, on construit une ligne brisée 14 à la surface de la sphère 16 en reliant les unes aux autres les extrémités des vecteurs par ordre de construction de ceux-ci.

La figure 43 est une représentation en mode non alterné ou géodésique. Pour l'obtenir, le signe de l'ordonnée de chaque vecteur d'observation est forcé à « + ».

Les mêmes variables sont illustrées sur la sphère à la figure 44 en mode dit alterné. Cette représentation en trois dimensions de type "pelure" correspond à la situation dans laquelle le signe de l'ordonnée de chaque vecteur d'observation est alterné à « + » ou « - ». Cette ligne forme la « pelure » précitée. Une fois obtenue, si la « pelure » du système complexe observé recouvre intégralement la sphère qui représente le système lui-même, la complétude du gisement des variables de supervision est à son tour vérifiée. Plus généralement, si la couverture n'est pas intégrale, on peut dire tout de même que plus la sphère est recouverte, plus le système de variables est pertinent.

D'une manière générale, l'affichage de type "pelure" permet d'évaluer la complétude du système de supervision, alors que l'affichage de type "géodésique" permet d'évaluer la répartition homogène des variables autour du système observé.

A ce stade, l'évaluation reste une démarche essentiellement visuelle.

Toutes les opérations de construction des représentations peuvent être avantageusement effectuées par le dispositif automatisé précité.

Bien entendu, on pourra apporter à l'invention de nombreuses modifications sans sortir du cadre de celle-ci.

On pourra effectuer la représentation offrant une supervision en trois dimensions au moyen du positionnement des vecteurs sur la sphère sans former des groupes constitués chacun par une des variables et d'autres des variables corrélées à cette dernière et suffisant pour déduire cette dernière, ni même associer, dans une base de données, à chacune des variables du groupe, une donnée indiquant au moins une identité des autres variables du groupe.

REVENDEICATIONS

1. Procédé pour organiser une base de données (4) comprenant des variables, telles que des indicateurs (I_1, I_2, I_3), relatives à un objet prenant plusieurs états, à
5 chaque variable correspondant des valeurs associées aux états respectifs, le procédé étant caractérisé en ce que :

- des moyens automatisés (2) identifient des corrélations entre au moins certaines des variables;

- les moyens (2) forment des groupes constitués chacun par une des variables (I_3) et d'autres (I_1, I_2) des variables corrélées à cette dernière et suffisant pour déduire
10 cette dernière; et

- les moyens (2) associent, dans une base de données (4), à chacune des variables du groupe, une donnée (M) indiquant au moins une identité des autres variables du groupe.

2. Procédé selon la revendication précédente dans lequel les moyens (2) construisent un graphe, tel qu'un réseau de Petri, du groupe, par exemple au moyen
15 d'un algorithme montant.

3. Procédé selon la revendication précédente dans lequel les moyens (2) codent le graphe pour produire la donnée (M).

4. Procédé selon au moins l'une quelconque des revendications précédentes dans lequel :

- on sélectionne une (I_0) des variables et on place un vecteur correspondant à la variable;

- on sélectionne une (I_1) des variables qui lui est le plus étroitement corrélée et on
25 place un vecteur correspondant à cette variable à une position représentant cette corrélation ;

- puis, autant de fois que nécessaire, on sélectionne une (I_i) des variables qui est le plus étroitement corrélée à celles correspondant à un vecteur déjà positionné et on place un vecteur correspondant à cette variable à une position représentant cette
30 corrélation,

les vecteurs étant positionnés avec leur origine au centre d'une sphère représentant l'objet et leur extrémité sur la sphère.

5. Procédé pour exploiter des variables, telles que des indicateurs (I_1, I_2, I_3), relatives à un objet prenant plusieurs états, à chaque variable correspondant des
35 valeurs associées aux états respectifs, le procédé étant caractérisé en ce que ;

- des moyens automatisés (2) obtiennent dans une base de données au moins une donnée (M) indiquant une identité (I_2 , I_3) de variables corrélées à une variable prédéterminée (I_1) et formant avec elle un groupe tel que les variables du groupe, sauf une (I_3), suffisent pour déduire cette dernière, les moyens identifiant ainsi les variables corrélées ; et

- les moyens (2) exploitent les identités des variables identifiées, par exemple les transmettent sur un réseau de télécommunication (8).

6. Procédé selon la revendication précédente dans lequel les moyens (2) identifient à partir de la donnée des relations de cause à effet entre les variables (I_1 , I_2 , I_3) du groupe.

7. Procédé selon l'une quelconque des revendications 4 et 5, dans lequel les moyens (2) commandent la mention sur une page web (6) des identités des variables identifiées.

8. Programme d'ordinateur caractérisé en ce qu'il comprend des instructions de code apte à commander la mise en œuvre d'un procédé selon l'une quelconque des revendications précédentes lorsqu'il est exécuté sur un ordinateur.

9. Dispositif (2) pour organiser une base de données (4) comprenant des variables (I_1 , I_2 , I_3), telles que des indicateurs, relatives à un objet prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs, le dispositif comprenant des moyens aptes:

- à identifier des corrélations entre au moins certaines des variables (I_1 , I_2 , I_3);
 - à former des groupes constitués chacun par une (I_3) des variables et d'autres (I_1 , I_2) des variables corrélées à cette dernière et suffisant pour déduire cette dernière;
- et

- à associer, dans une base de données (4), à chacune des variables du groupe, une donnée (M) indiquant au moins une identité des autres variables du groupe.

10. Dispositif (2) d'exploitation de variables (I_1 , I_2 , I_3), telles que des indicateurs, relatives à un objet prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs, le dispositif, éventuellement selon la revendication précédente, étant caractérisé en ce qu'il comprend des moyens aptes ;

- à obtenir dans une base de données (4) au moins une donnée (M) indiquant une identité (I_2 , I_3) de variables corrélées à une variable prédéterminée (I_1) et formant avec elle un groupe tel que les variables du groupe, sauf une (I_3), suffisent pour déduire cette dernière, les moyens étant aptes ainsi à identifier les variables corrélées ; et

- à exploiter les identités des variables identifiées, par exemple à les transmettre sur un réseau de télécommunication (8).

11. Base de données (4) caractérisée en ce qu'elle comprend :

5 - des variables (I_1, I_2, I_3) relatives à un objet prenant plusieurs états, à chaque variable correspondant des valeurs associées aux états respectifs;

- au moins une donnée (M) en relation avec au moins une des variables et indiquant une identité de certaines des variables corrélées avec elle et formant avec elle un groupe tel que les variables du groupe, sauf une (I_3), suffisent pour déduire cette dernière.

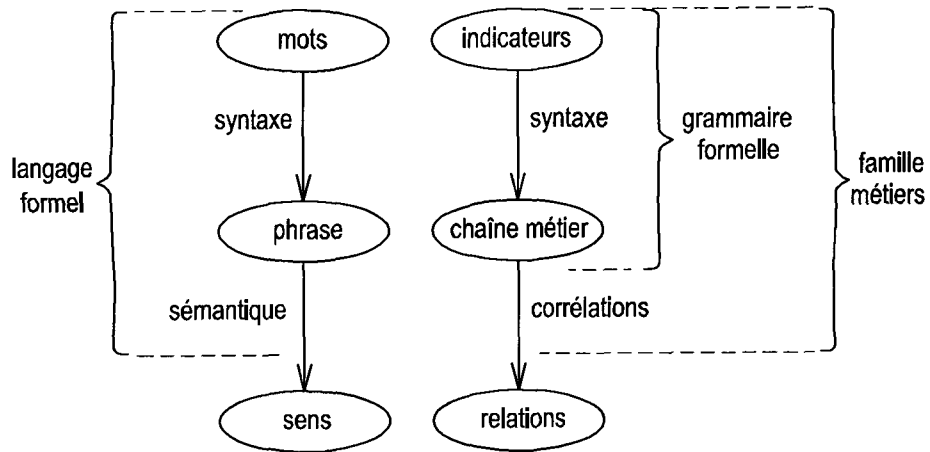


Fig. 1

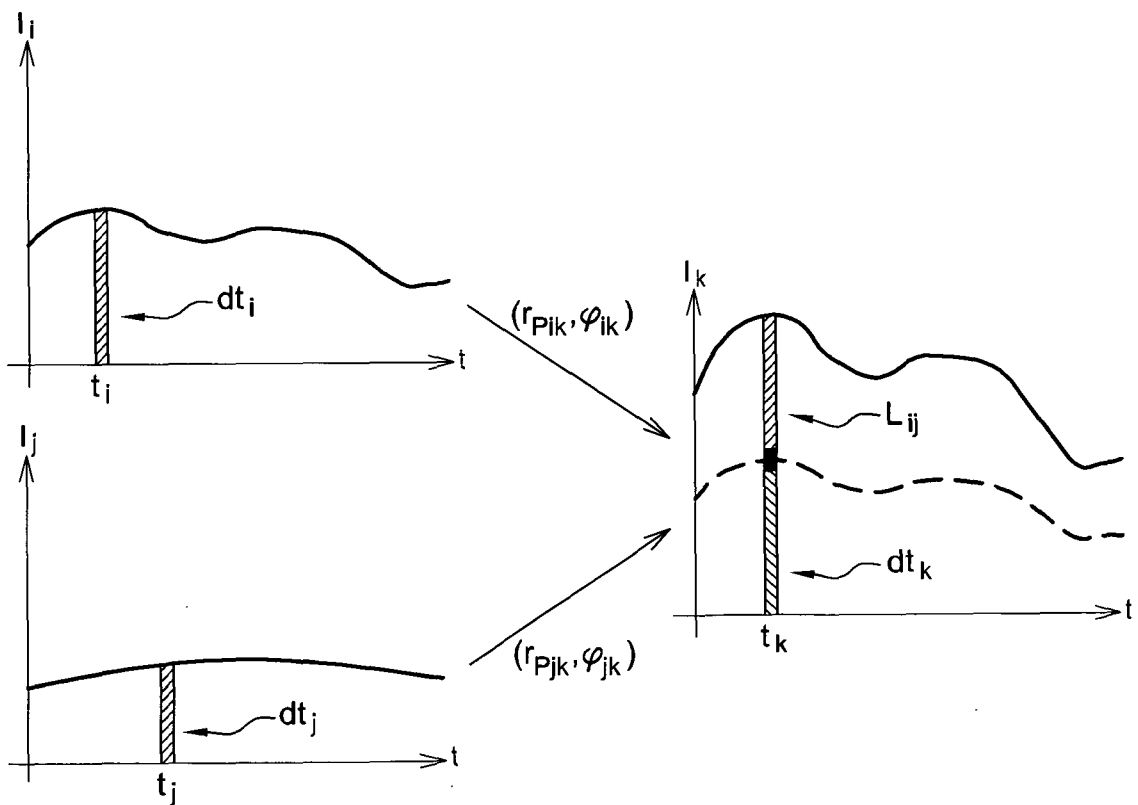


Fig. 2

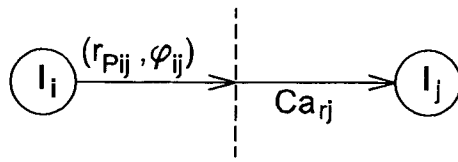


Fig. 3

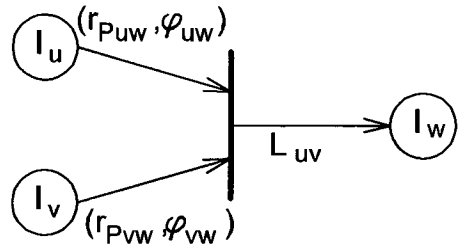


Fig. 4

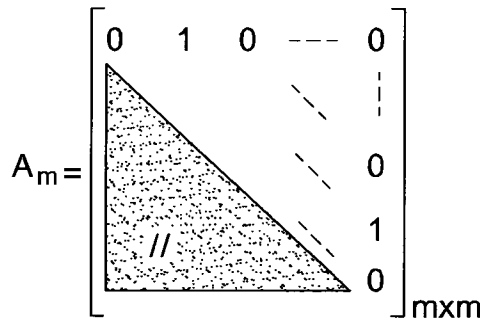


Fig. 5

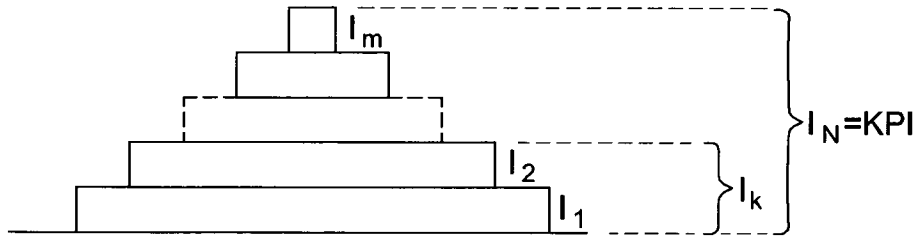


Fig. 6

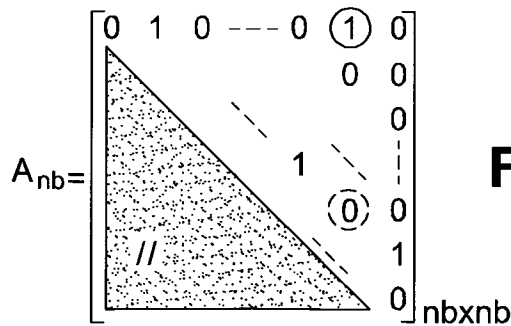


Fig. 7

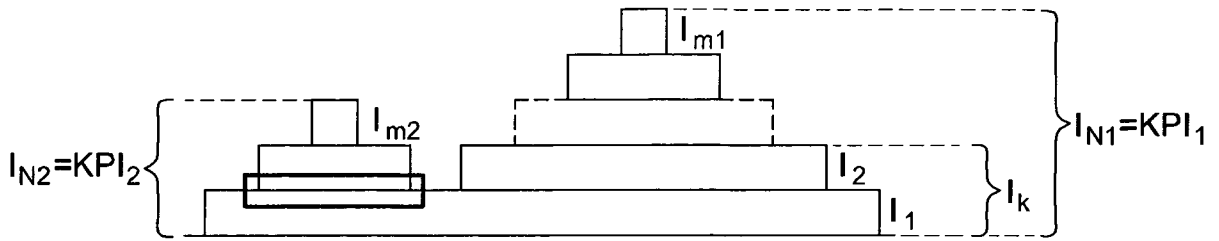


Fig. 8

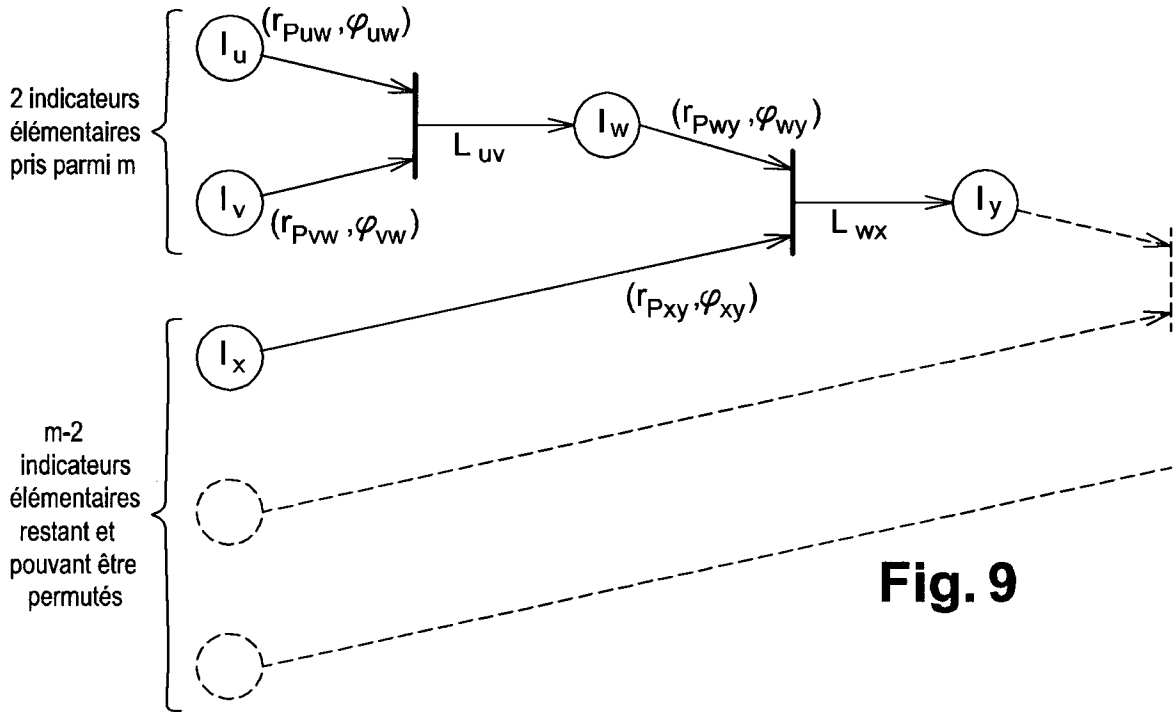


Fig. 9

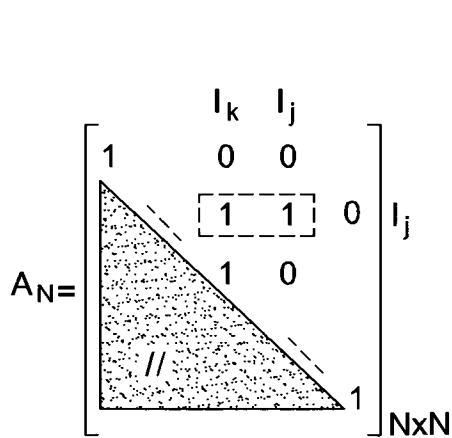


Fig. 10

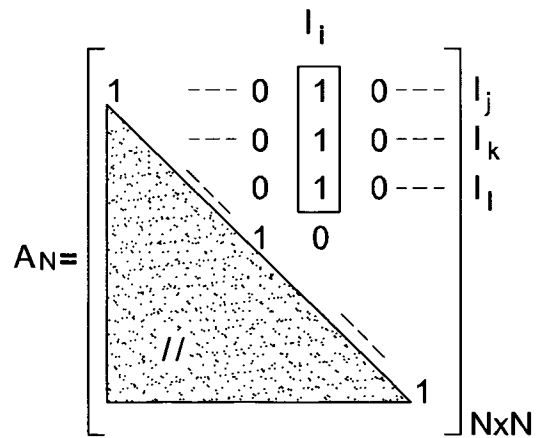


Fig. 11

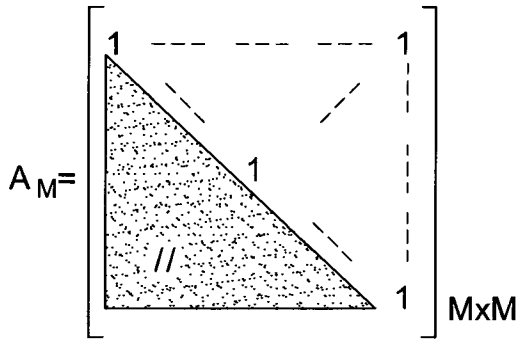


Fig. 12

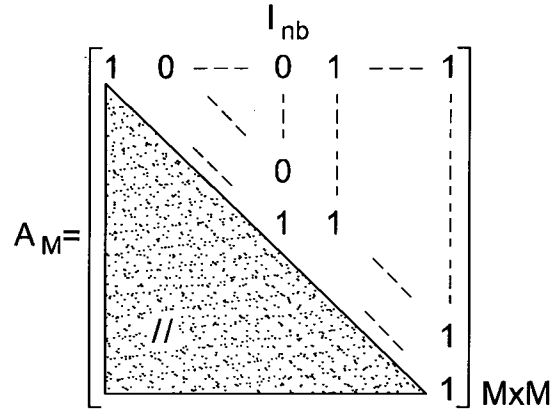


Fig. 13

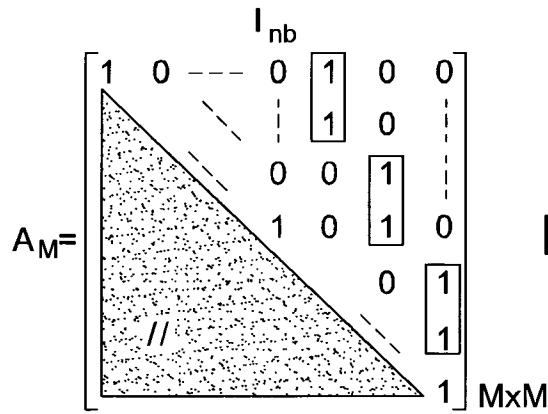


Fig. 14

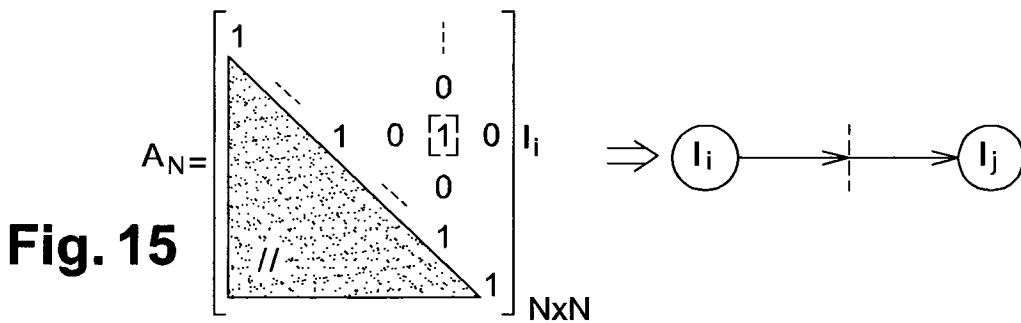


Fig. 15

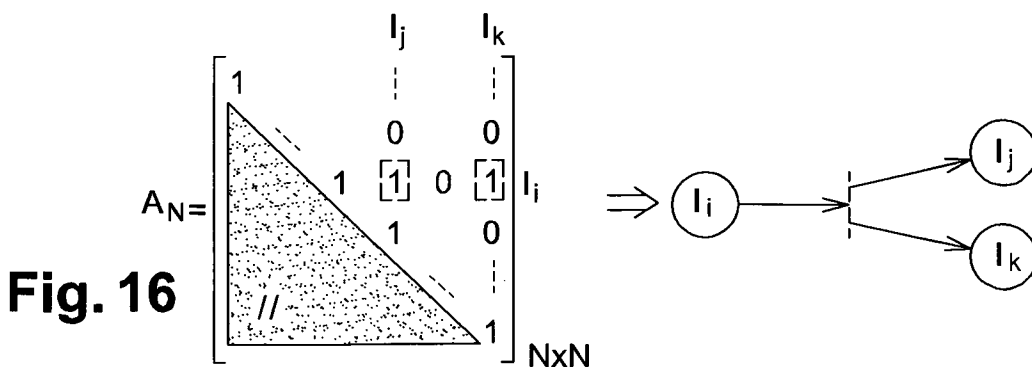


Fig. 16

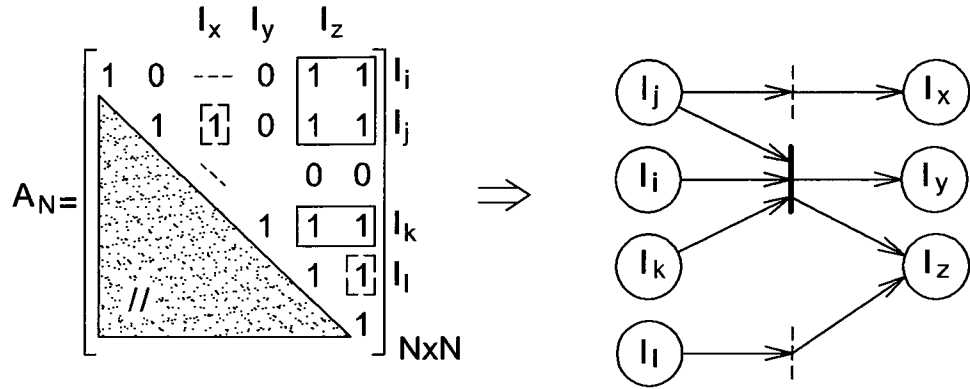


Fig. 21

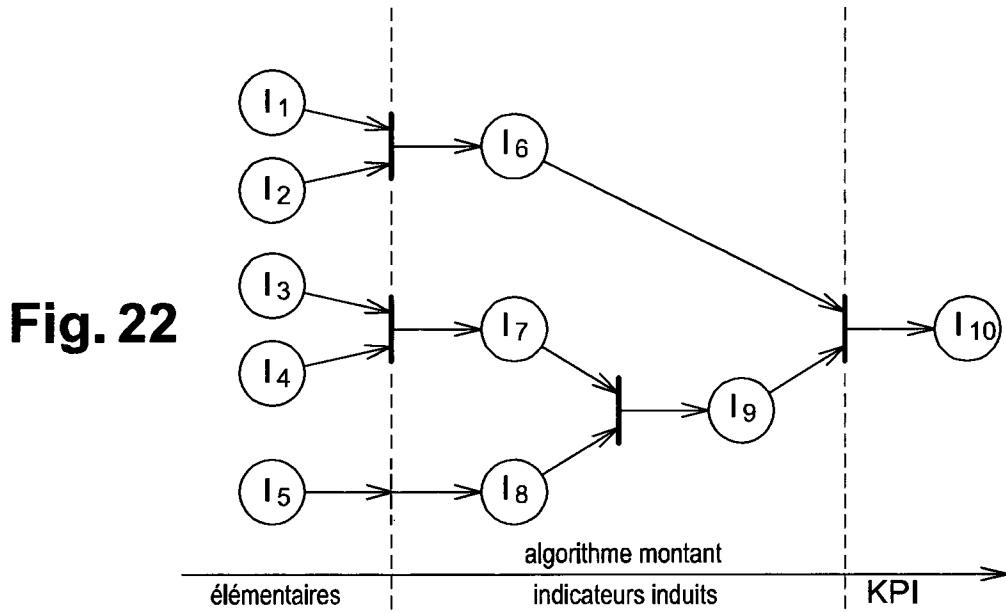


Fig. 22

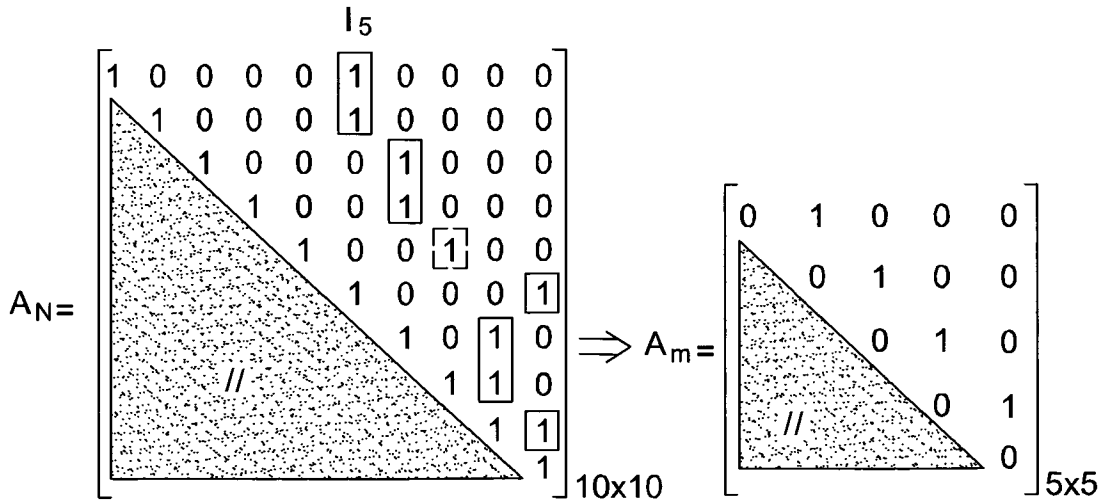
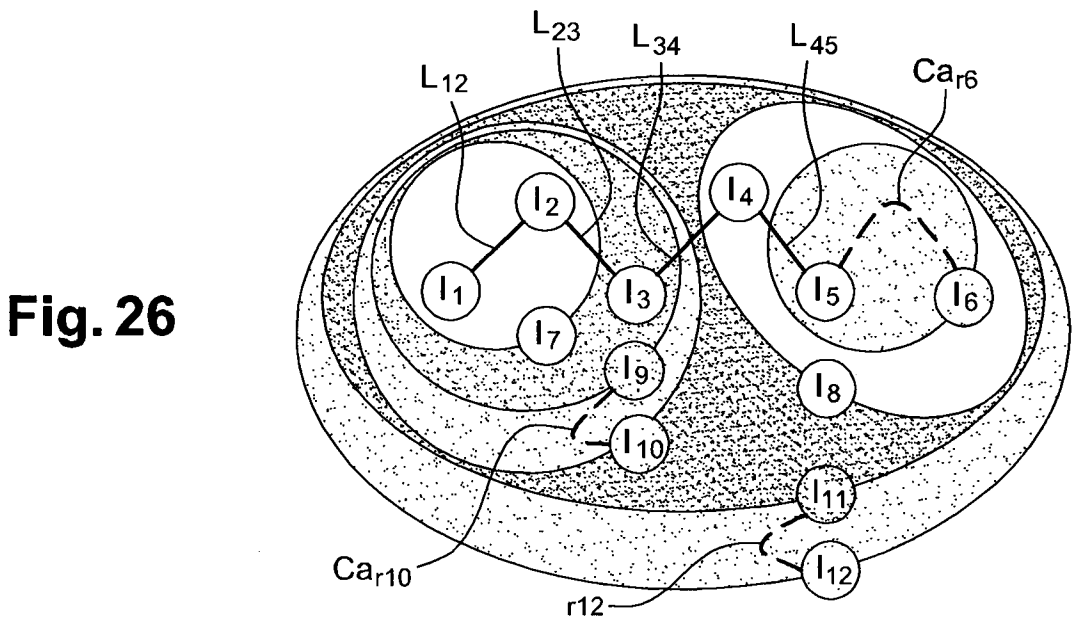
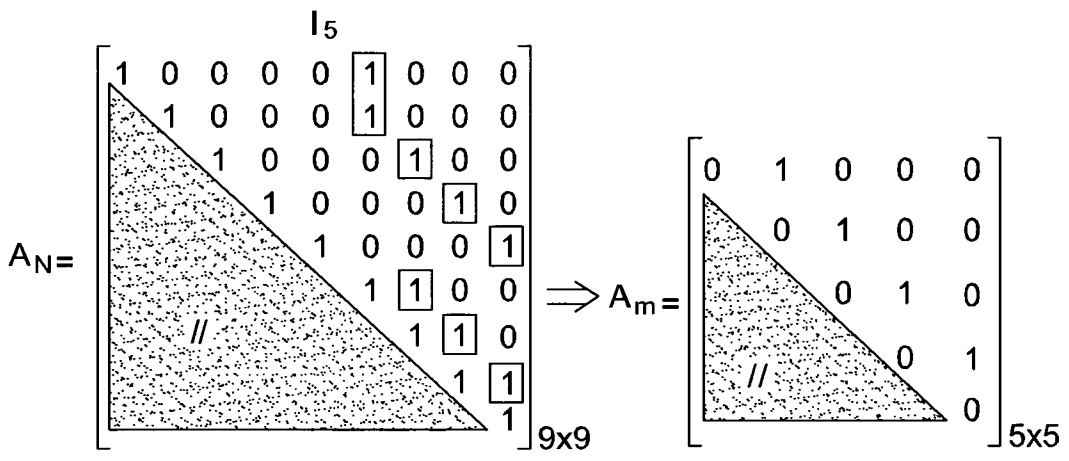
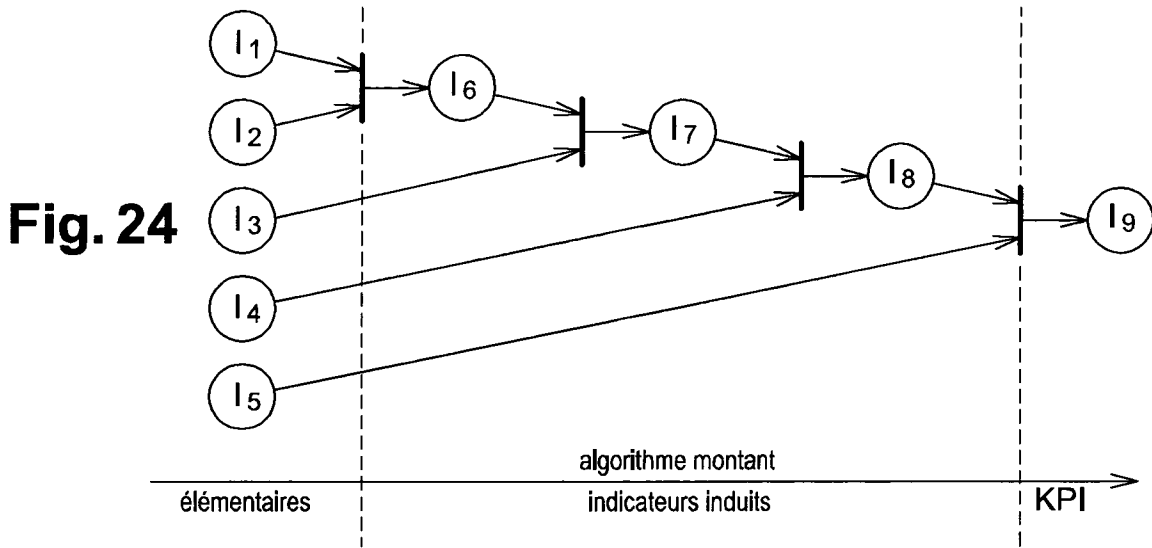


Fig. 23



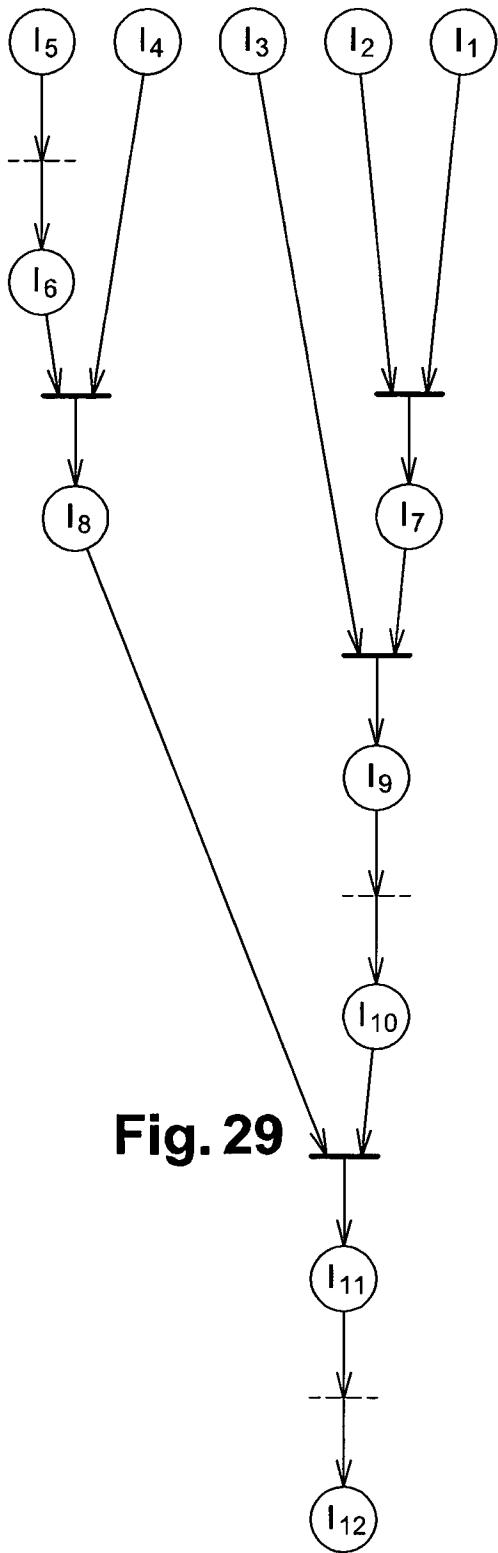


Fig. 29

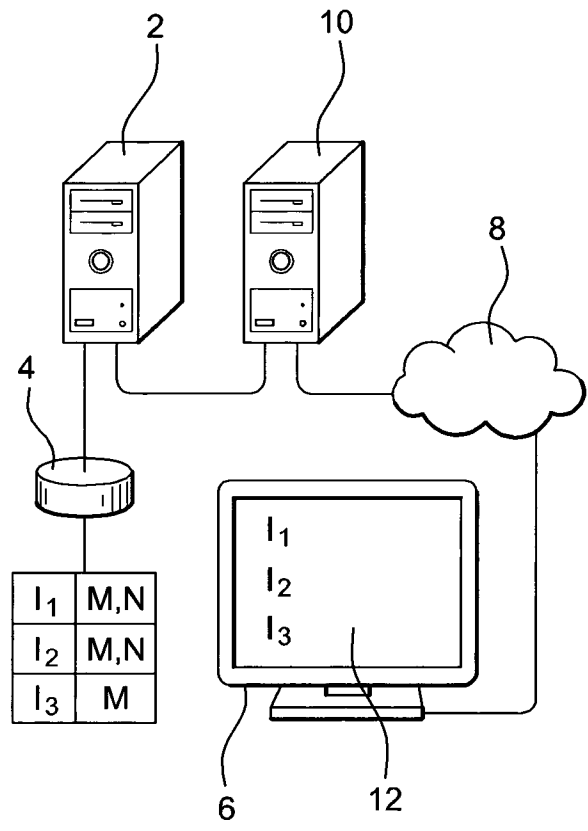


Fig. 30

Fig. 31

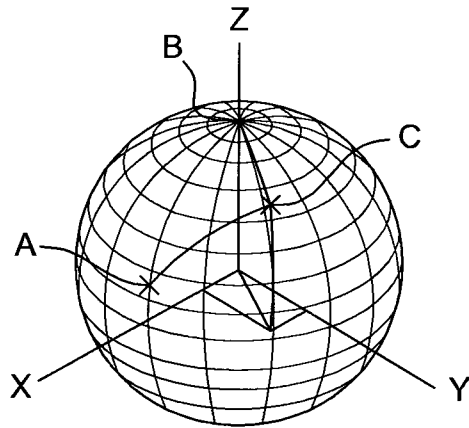


Fig. 32

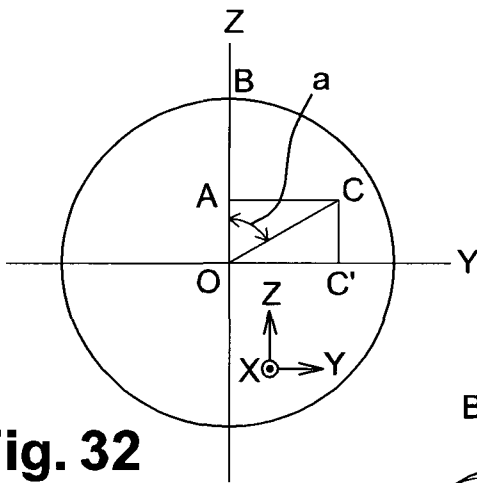


Fig. 33

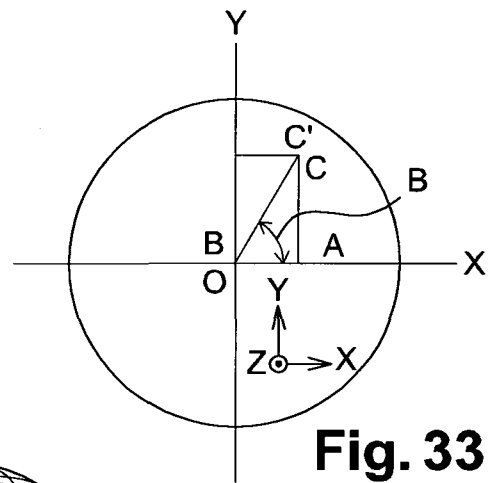


Fig. 34

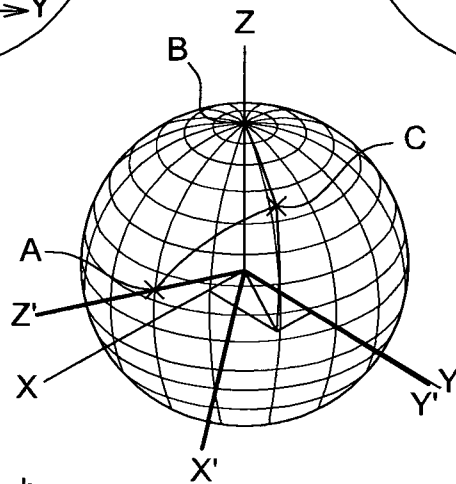


Fig. 35

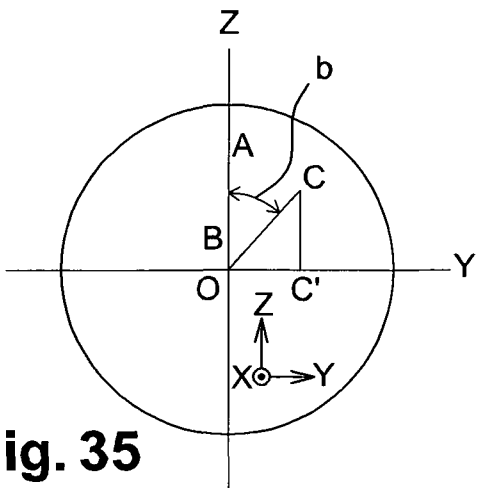
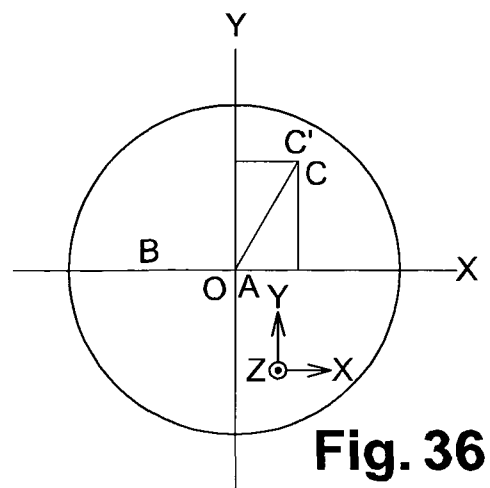


Fig. 36



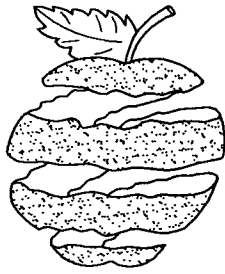


Fig. 37

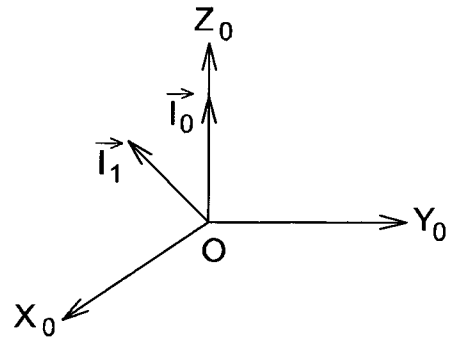


Fig. 38

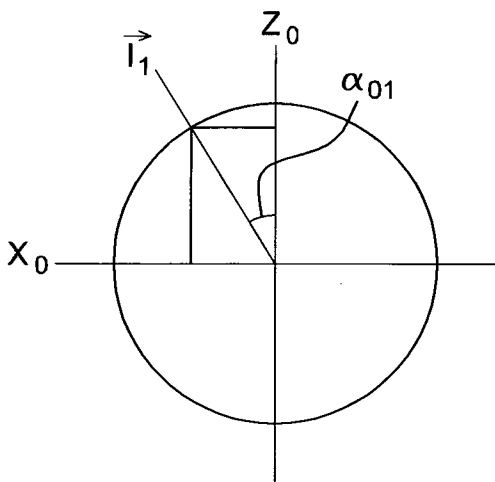


Fig. 39

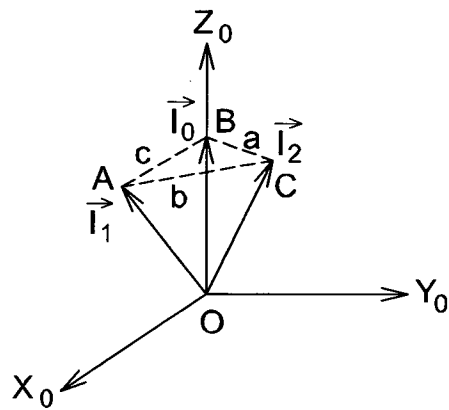


Fig. 40

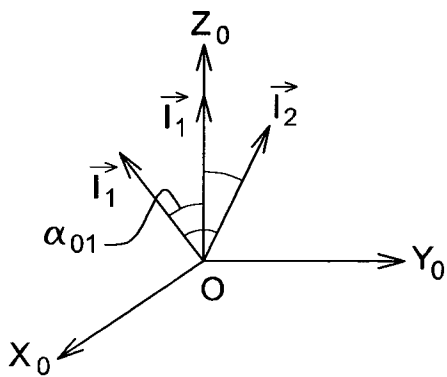


Fig. 41

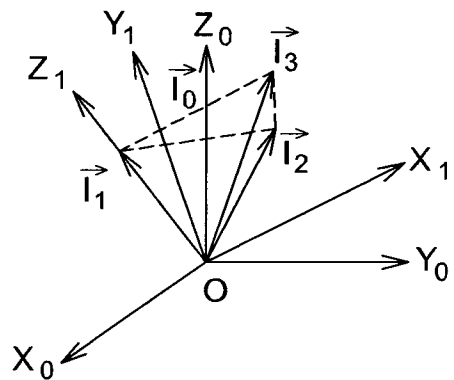


Fig. 42

