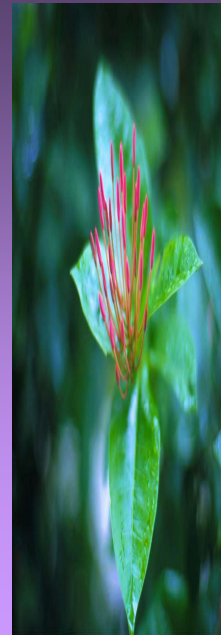


KABA
Langage d'indexation
matière
Chantiers de travail

Maria Nasifowska

Journée nationale d'information
RAMEAU
Bnf, 30 mai 2007



Langage d'indexation KABA

Chantiers de travail

- 1. Construction du vocabulaire KABA**
- 2. Application du langage KABA**
- 3. Évolution du système KABA**

... en coopération avec le bureau RAMEAU

Plan de communication:

1. La construction du vocabulaire: les origines du langage KABA; la méthode de la construction du vocabulaire
2. L'application du langage: le réseau des bibliothèques NUKAT; KABA en chiffres
3. L'évolution du système KABA, tendances principales: le typage des vedettes d'autorité et des éléments de la vedette matière construite (VMC); le travail sur le vocabulaire

Tous ces sujet on verra à la lumière de la coopération avec le bureau RAMEAU.

Construction du vocabulaire KABA Début

- ◆ Démarrage du projet de Mme Teresa Głowacka en 1992
- ◆ Objectif – indexation matière des fonds des bibliothèques universitaires et scientifiques, dans les catalogues informatisés
- ◆ Choix du modèle – RAMEAU
- ◆ Coopération depuis le début avec Mme Barbara Kotalska et le Centre RAMEAU

1. Le langage d’indexation KABA est construit depuis le 1992.
2. L’objectif dans cet entreprise était l’indexation des documents dans les bibliothèques universitaires et scientifiques de Pologne
3. Le choix a été fait entre: le langage d’indexation matière géré par la Bibliothèque nationale de Pologne, le langage américain – LCSH, le langage RAMEAU. La decision était pour RAMEAU, le langage d’indexation de la BN de Pologne étant trop générale et préparé surtout pour l’usage dans les bibliothèques publiques. L’argument pour RAMEAU, et non LCSH, c’ était la proximité des réalités, de la culture, etc.

Construction du vocabulaire KABA Continuation

- ◆ Vedette d'autorité créée:
 - pour l'indexation d'un document et non dans l'abstrait,
 - en notices d'autorité – système des relations entre vedettes,
 - à la base de la notice d'une vedette RAMEAU ayant le sens identique ou le plus proche,
 - signalant l'équivalent de RAMEAU et la source méthodologique – RAMEAU.

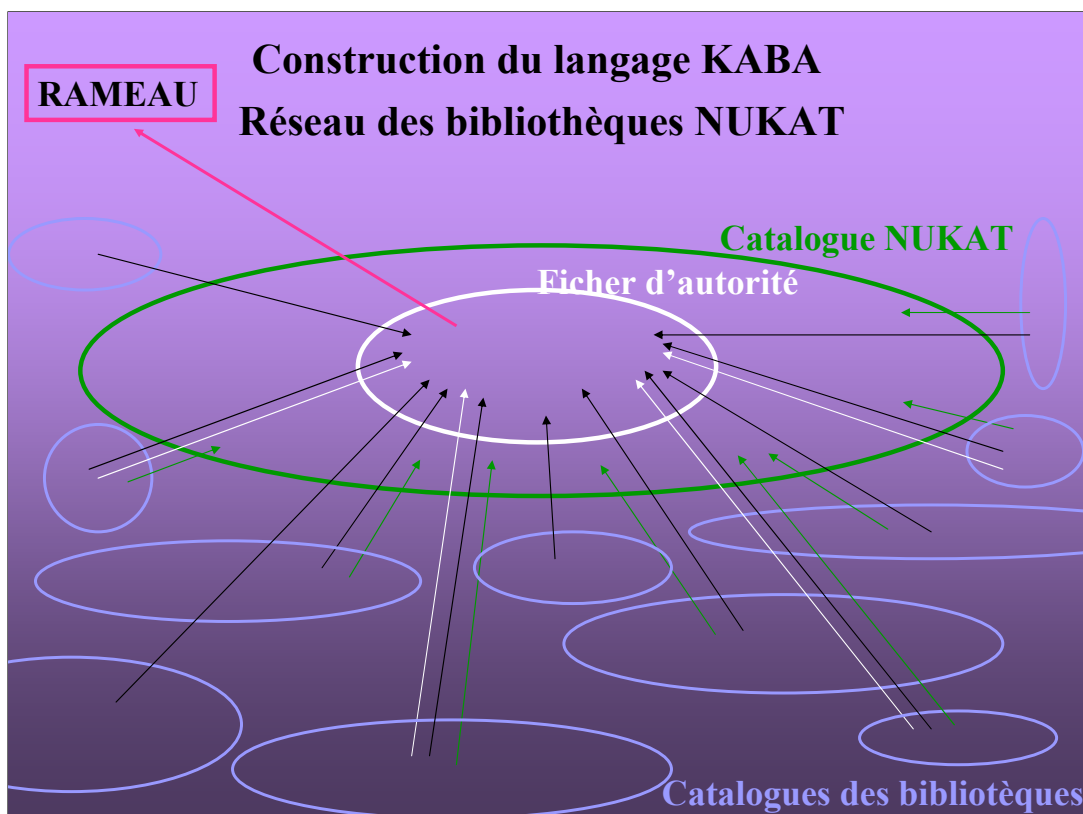
Quelques règles de base de la construction du vocabulaire KABA - Les vedettes d'autorité KABA sont construits:

- pour les besoins de l'indexation courante des documents dans le catalogue collectif NUKAT
- en notice d'autorité – chaque nouvelle vedette est liée avec les vedettes existantes dans le vocabulaire KABA
- sur le modèle de la notice de la vedette RAMEAU qui a le sens identique ou le plus proche de la vedette en construction pour KABA. On essaye de signaler dans la notice les mêmes liens que les liens de la notice RAMEAU. Cependant la structure finale de la notice d'autorité KABA dépend aussi de la sémantique du terme polonais et du format de travail (MARC 21), - voir la diapo suivant.
- dans les notices des vedettes d'autorité KABA on signale la vedette RAMEAU et la vedette LCSH dont le sens et le même ou le plus proche de la vedette KABA. On ajoute en zone des sources le mot „RAMEAU” pour signaler que le modèle pour la construction de la notice était la notice du langage RAMEAU (on appelle ce type de source la source méthodologique).

Construction du vocabulaire KABA Continuation

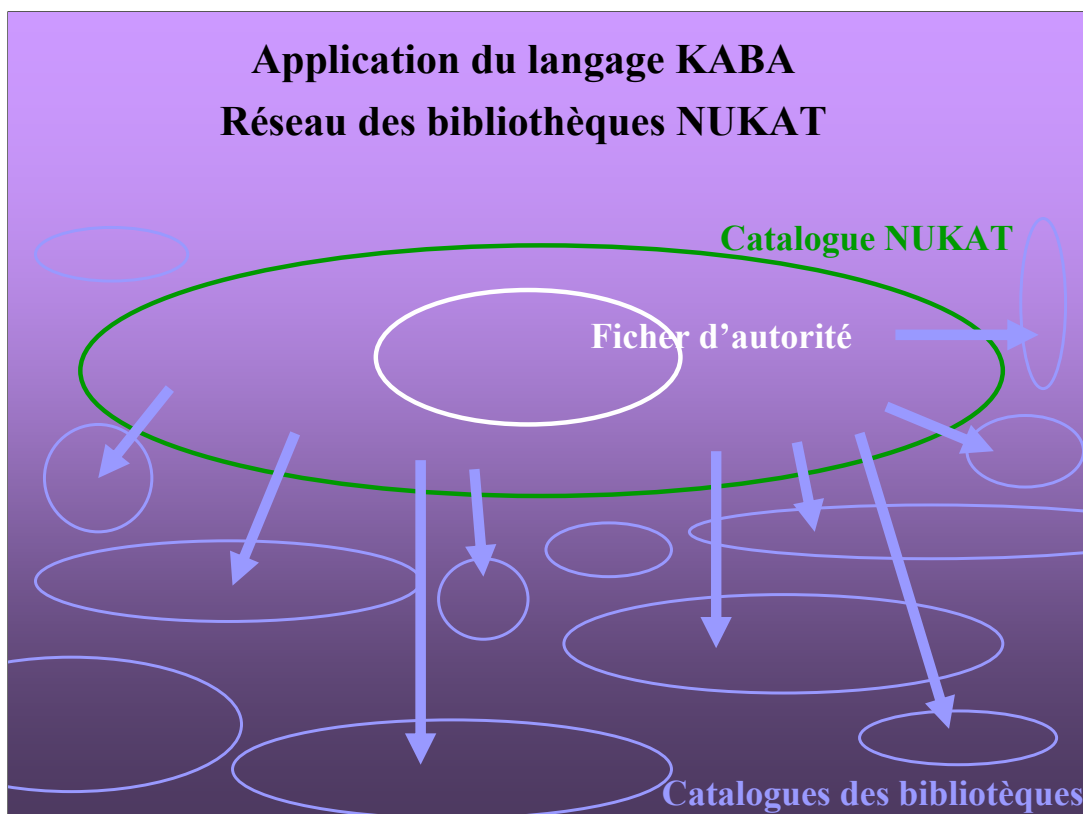
- ◆ **Autonomie des vedettes d'autorité KABA par rapport à la vedette RAMEAU:**
 - différences des réalités polonaises
 - différences sémantiques des termes
 - différences des fonds des bibliothèques
 - différences du contexte – format des données, système intégré, type de catalogue

Bien que les notices d'autorité de vedettes KABA sont construites sur le modèle de la notice RAMEAU, pour des raisons suivantes elle peut avoir finalement la structure différente.



Méthode de la construction du langage KABA en réseau des bibliothèques NUKAT:

1. La bibliothèque qui indexe un document et constate le manque d'une vedette d'autorité KABA, construit la notice d'autorité de la vedette KABA dans le fichier d'autorité NUKAT (flèches blanches). La vedette d'autorité une fois entrée au fichier peut être utilisée dans l'indexation. Parmi environ 80 bibliothèques collaborantes à NUKAT, une partie seulement peut construire les vedettes d'autorité KABA. Dans le catalogue NUKAT il y a deux autres langages d'indexation matière (langage de la Bibliothèque nationale de Pologne et le langage médicale - MeSH).
2. (Flèches noires) – Toutes les bibliothèques de NUKAT construisent les notices d'autorité auteurs et titres – deuxième partie du fichier d'autorité NUKAT
3. Flèches vertes indiquent la construction des vedettes matière construites (VMC) du langage KABA. Cette étape appartient déjà à la méthode de l'application du langage KABA, voir la diapo suivante....



Méthode de l'application du langage - Réseau des bibliothèques NUKAT

Flèches bleues signifient le transfert des notices bibliographiques complètes. Dans la méthode de l'application du langage KABA il n'a pratiquement pas de coopération avec bureau RAMEAU. KABA est un langage d'un catalogue. Le catalogue qui est un catalogue collectif pour la majorité des grandes bibliothèques en Pologne. Pour cela il est nécessaire de construire une seule méthode de l'application des vedettes d'autorités, ainsi que des vedettes matières en générale.

Comme dans le catalogue NUKAT il existe deux autres langages d'indexation matière, l'usage de KABA n'est pas obligatoires pour la bibliothèque appartenante au réseau NUKAT – c'est un choix volontaire. Cependant le choix une fois fait on est obligée de respecter une seule méthode de l'application de KABA. Les chiffres illustrant cette situation, voir la diapo suivante...

Application du langage KABA

Quelques chiffres

- ◆ NUKAT - 80 bibliothèques
 - ◆ 62 - utilisent KABA
 - ◆ 21 - construisent KABA

- ◆ Le langage d'indexation KABA
 - ◆ 252658 - vedettes matière d'autorité (VMA)
 - ◆ 1045754 - vedettes matière construites (VMC)

1. Le réseau du catalogue collectif NUKAT compte 80 bibliothèques parmi lesquelles il y a des bibliothèques universitaires, polytechniques, scientifiques et aussi quelques bibliothèques d'information publiques. Les bibliothèques qui entrent en coopération dans le réseau NUKAT choisissent le langage d'indexation matière. Si le choix est pour KABA, la bibliothèque passe par le système de formation en deux niveaux. Le premier niveau c'est l'application du langage dans l'indexation, le niveau suivant c'est la construction des vedettes d'autorité KABA.

2. Le langage KABA c'est 252658 des VMA et 1045754 des VMC. La différence entre VMA et VMC pose un certain nombre des problèmes au niveau de la construction du langage KABA. Récemment nous avons réussi à établir la distinction, c'est-à-dire les critères selon lesquels on décide à l'étape de l'indexation du document si la vedette est une vedette d'autorité ou la VMC.

Évolution du système KABA Tendances principales

- 1. Typage des noms des sujets**
- 2. Typage des éléments des noms des sujets**
- 3. Travail sur le vocabulaire**

Actuellement à part le travail quotidien sur le vocabulaire, le bureau KABA s'occupe des sujets liés à l'évolution du langage KABA. Les principales tendances de notre réflexion se placent sur trois niveaux:

1. Le niveau du fichier d'autorité – c'est le typage des noms de sujets.
2. Le niveau de la vedette matière construite (VMC) – le typage des éléments de la VMC
3. Le dernier dans cette liste c'est ce qui paraît le plus important – le travail sur la terminologie correcte et actuelle du langage KABA.

Dans la suite on verra en quoi consiste le travail sur ces trois niveaux.

Évolution du système KABA Typage des noms des sujets

Langage	Vedette d'autorité (VA)
KABA	(CO) Yellowstone National Park (CO) British Library. Rękopis. Cotton Tiberius B.VIII (NC) Rāma (bóstwo hinduskie) (NC) Literatura fantastyczna; (F) Literatura fantastyczna
RAMEAU	(CO) Yellowstone National Park (États-Unis) (CO) British Library -- Manuscrits. Ms. Cotton Tiberius B. VIII (NP) Rāma (divinité hindoue) (NC) Littérature fantastique
LCSH	(NG) Yellowstone National Park (CO) British Library. Manuscript. Cotton Tiberius B. VIII (NC) Rāma (Hindu deity) (NC) Fantasy literature; (F) Fantasy literature [GSAFD]

Le typage des vedettes d'autorité joue le rôle principal dans le processus de la construction du vocabulaire. Il est important aussi dans l'application du vocabulaire – la façon de construire des VMC dépend du typage (le mécanisme des listes des subdivisions). En comparant le typage des vedettes d'autorité KABA, RAMEAU et LCSH on voit les différences, ce qui en conséquence génère les différences dans l'indexation matière.

Dans le processus de la construction du vocabulaire KABA la source principale c'est le langage RAMEAU – cette principe est valide aussi pour le typage des vedettes d'autorité. Ce fait provoque souvent des problèmes parce qu'en réalité il est impossible de tout transmettre du format INTERMARC au format MARC 21. En plus le format MARC 21 qui est un format de travail pour le langage KABA, impose en quelque sens le typage des vedettes d'autorité (voir le tableau de typage dans le site internet de la Bibliothèque du Congrès, format MARC 21 autorité). Finalement dans le travail sur les vedettes KABA souvent on décide de changer le typage et de profiter des règles du format MARC 21 (exemple des vedettes pour les manuscrits).

Le typage des vedettes d'autorité consiste à répondre à la question sur la nature ou fonction d'une entité (le problème si c'est la nature ou la fonction est un autre grand problème du typage). La réponse à cette question est ensuite à coder dans le format de travail.

Le souci de bien coder la nature des entités (ce qui joue le rôle essentiel dans la présentation des données dans le catalogue) nous a mené à prendre la décision sur l'application d'un nouveau type de vedettes d'autorité – vedettes de formes. Ces vedettes ne désignent pas le sujet de l'oeuvre, mais la forme que représente l'oeuvre.

Les problèmes de typage, et souvent le typage différent dans le cadre d'un même type de nom selon si on prend pour modèle la vedette RAMEAU ou LCSH, nous ont fait construire le tableau du typage des vedettes d'autorité (voir la diapo suivante...).

CO – nom de collectivité; NC – nom commun; NG – nom géographique; NP – nom de personne;
F – forme

GSAFD - Guidelines on Subject Access to Individual Works of Fiction , Drama, Etc.

Évolution du système KABA Tableau de typage des noms des sujets

Nom de sujet	Fichier	Étiquette	Vedette d'autorité - exemple
dieux, deesses	FS	150	Apollon (bóstwo greckie)
églises (confession)	FAT	110	Kościół Ewangelicko-Reformowany w Polsce
festivals	FAT	111	Festiwal Mozartowski (16 ; 2006 ; Warszawa)
films	FS	130	Matrix (film)
groupes de musique	FAT	110	Rolling Stones
jardins botaniques	FAT	110	Uniwersytet Warszawski. Ogród Botaniczny
villes anciennes	FS	151	Troja (miasto dawne)

Le diapo montre un fragment du tableau de typage des vedettes d'autorité KABA.

Le tableau, qui notamment est encore en construction, va constituer une aide pour la construction des vedettes d'autorité.

Dans la première colonne on va noter les noms des entités qui peuvent être les sujets des documents. Il n'y a ici que les noms qui sont difficiles à typer, on ne va pas y mettre les noms qui ne posent pas des problèmes comme par exemple nom des plantes (qui sont toujours typés comme Non Commun (étiquette 150 dans le format MARC 21). La deuxième colonne indique auquel fichier appartient le nom de sujet - fichier „sujet” KABA (FS) ou fichier „auteurs, titres” (FAT).

La colonne suivante indique l'étiquette de la vedette d'autorité. Enfin la dernière colonne donne l'exemple de la vedette d'autorité. Il faut noter que le type d'étiquette n'est pas lié au fichier, c'est-à-dire il arrive qu'on construise un type des vedettes d'autorité dans le fichier sujet ayant l'étiquette de la vedette propre au fichier „auteurs, titres” (110 – nom de collectivité; 111 – nom de collectivité (événement); 130 – titre uniforme). Le fichier „sujet” ne contient pas seulement les vedettes avec ces étiquettes typiquement matière (150 – nom commun; 151 – nom géographique).

Évolution du système KABA Typage des éléments des noms des sujets

Nom de sujet -

- vedette matière construite (VMC):

- ◆ **ordre des éléments**
- ◆ **definition des éléments**
- ◆ **typage de éléments**

Sujet du document peut aussi être exprimé par la vedettes matière construite (VMC), composée de la tête de vedette et des subdivisions. L' établissement du typage des vedettes d'autorité concerne donc la première partie de la VMC. Le problème du typage – la question sur la nature ou fonction (ce qu'on va voir dans la suite) est aussi intéressante à poser dans le cas des éléments de la VMC. La réflexion sur le typage des éléments de la VMC se compose de trois étapes: l'ordre, la définition et le typage éléments.

Évolution du système KABA Typage des éléments des noms des sujets

L'ordre des éléments de la VMC, tendances à :

- ♦ retour à l'ordre original
- ♦ élimination des exceptions

TV SS SG SCH SF

VF SG SCH SF

L'ordre des éléments de la VMC est fixe, cependant il existe des exceptions, des règles particulières qui font que l'ordre est différent. Deux tendances dans le travail quotidiens devrait nous aider à simplifier la structure des noms de sujets – les VMC: c'est le retour à l'ordre original. Il y a ici deux types des structure: celle typiquement de sujet – TV—SS—SG—SCH—SF et la deuxième: VF—SG—SCH—SF. La deuxième tendance c'est l'élimination des exceptions des ces deux type des structure.

TV – tête de vedette; SS – subdivision de sujet; SG – subdivision géographique; SCH - subdivision chronologique; SF - subdivision de forme; SV – vedette de forme

Évolution du système KABA Typage des éléments des noms des sujets

Definition des éléments de la VMC selon:

- ◆ nature
- ◆ fonction

Alchimie -- France -- Paris -- 14e siècle -- Sources

Alchimie -- Histoire -- Congrès

Français (langue) -- 1300-1500 (moyen français)

Français (langue) -- Langage parlé -- Canada -- 20e siècle

Poésie française -- Avant 1500

Musique vocale -- 20e siècle

Littérature américaine -- Auteurs noirs américains

Après l'ordre des éléments de la VMC il y a la question sur la de définition des éléments. Il s'agit ici de répondre à la question sur ce qu'exprime l'élément de la VMC. La réponse dépend de point de vu – si on prend un élément isolé, par exemple la subdivision „1300-1500 (moyen français)” on constate qu'elle exprime la période, le temps. On le définit donc comme la subdivision chronologique. Cependant quand on regarde un élément dans le contexte on peut essayer de poser la question sur la fonction qu'elle joue dans ce contexte. Autrement dit – quelle est la relation de cet élément envers l'élément qui le précède. Dans le cas de la subdivision „1300-1500 (moyen français)” on voit que sa fonction est autre que celle de la subdivision „14e siècle” qui situe „Alchimie” dans le temps, alchimie qui existait aussi bien au 14e siècle qu'au 19e siècle, c'était toujours le même phénomène, qu'on nome „alchimie”. Dans le cas de la langue française, la différence entre le moyen français et le français contemporain est telle qu'on peut dire que ce sont deux langues. La subdivision „1300-1500 (moyen français)” ne situe pas le sujet dans le temps mais précise le sujet – c'est le rôle de la subdivision de sujet. Cette question reste ouverte – dans le typage des subdivisions dans le langage KABA comme RAMEAU on prend en considération la nature plutôt que la fonction des éléments.

La même question est à poser par rapport à un certain type des têtes de vedette – les vedettes dont la fonction est d'exprimer non le sujet mais la forme du document indexée, ou plutôt du contenu de ce document. Ces vedettes, typées comme vedettes de forme vont être codées autrement que les vedettes typiquement de sujet. Voir la diapositive suivante.

Évolution du système KABA Typage des éléments des noms des sujets

Typage de éléments de la VMC

- ◆ cogade au format selon le typage

TV Alchimie SG France SG Paris SCH 14e siècle SF Sources

TV Alchimie SS Histoire SF Congrès

TV Français (langue) ?? 1300-1500 (moyen français)

TV Français (langue) SS Langage parlé SG Canada SCH 20e siècle

VF Poésie française ?? Avant 1500

VF Musique vocale ?? 20e siècle

VF Littérature américaine SF Auteurs noirs américains

Le typage des éléments s'exprime dans le codage au format de travail, ce qui est un moyen de communication avec la machine. Le format est le langage lisible par la machine – le cogade permet les opérations automatiques sur les données. Dans le travail sur le langage KABA le cogade correcte au format MRC 21 prend une place importante. Dans l'histoire de ce langage souvent on prenait les décisions pas conformes à ce format MARC 21 – c'est parce que le modèle principal pour KABA c'est RAMEAU. Le format de travail des RAMEAU est INTERMARC. Ces deux formats ne sont pas compatibles – je veux dire qu'il y a des informations codées dans INTERMARC qu'on ne peut pas exprimer dans MARC 21. L'autonomie ces deux langages s'impose aussi à cause de ce problème.

Le problème du codage (et d'abord du typage) des éléments comme la subdivision „1300-1500 (moyen français)” reste ouverte ...

TV – tête de vedette; SS – subdivision de sujet; SG – subdivision géographique; SCH - subdivision chronologique; SF - subdivision de forme; VF – vedette de forme

Évolution du système KABA Travail sur le vocabulaire

- ◆ Révision de la terminologie
 - ◆ mise à jour des termes spécialisés
 - ◆ développement de la terminologie „locale”
- ◆ Simplification des structures – VA,VMC
- ◆ **Relation: RAMEAU - KABA**
 - ◆ **collaboration effective au niveau méthodologique**
 - ◆ **autonomie au niveau terminologique**

Chantier de travail le plus exploré c'est le travail sur le vocabulaire. En créant des vedettes m'autorité et en modifiant des vedettes existantes dans KABA nous mettons beaucoup d'attention sur l'actualité des termes spécialisés. Souvent il est nécessaire de développer la terminologie, que j'appelle ici „locale” – je veux dire autonome des structures qu'on trouve dans RAMEAU.

Dans le travail sur le vocabulaire il nous arrive de réaliser le projet dont j'ai déjà parlé – la simplification des structures des vedettes.

Enfin un mot à propos de la coopération avec le bureau RAMEAU – cette coopération qui nous est très utile concerne actuellement surtout le niveau méthodologique. Le travail sur la terminologie reste une question locale.

Merci de votre attention!

