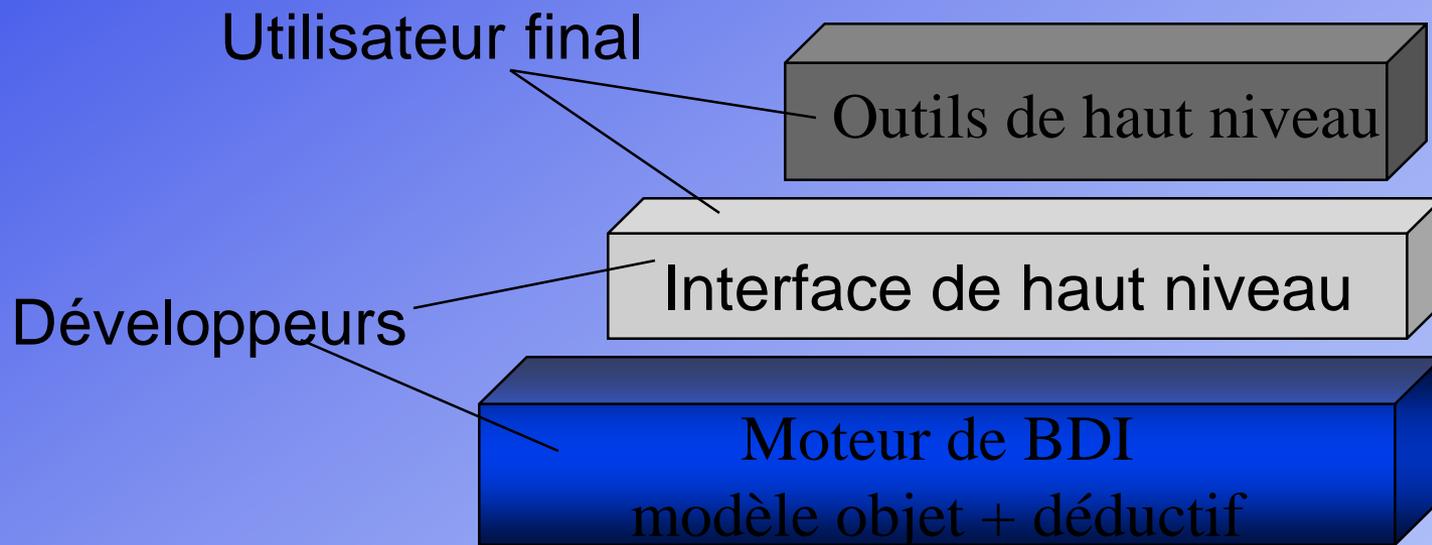


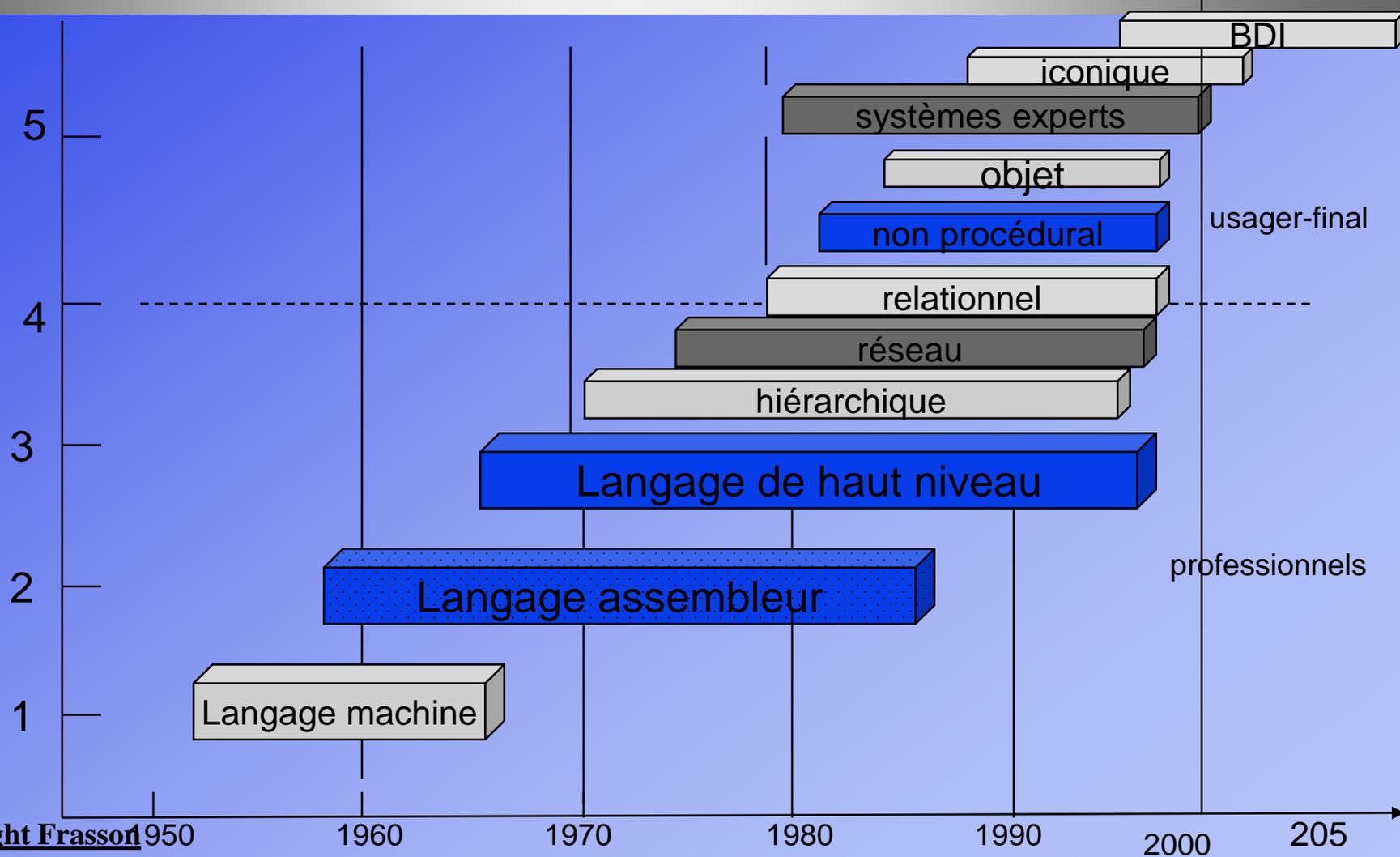
Bases de données intelligentes

- Nouveaux types de bases de données incorporant plusieurs fonctionnalités



Combinaison de hypermedia, QBE, inférence, O-O

L'évolution des bases de données



Principes généraux

- Bases de données qui incorporent des composants intelligents
 - Aide pour les opérations de recherche
 - Méthodes de représentation de connaissances
 - Réseaux de neurones connexionnistes (association avec des informations similaires connues)
 - Compréhension des problèmes

Caractéristiques des BD intelligentes

- Feedback : s'adapte au retour de l'utilisateur
 - Exemples de ce que l'utilisateur cherche ou ne cherche pas
 - Profils de recherche
- Interface
 - Arrangement de fenêtres détaillant les recherches précédentes de l'utilisateur
 - Pointage direct des termes

Aide à l'utilisateur

- Fournie de plusieurs façons et à plusieurs niveaux
- Expertise pour usagers spécialisés
- Suggestion de solutions
- Analyse des commentaires de l'utilisateur
- Reformulation des requêtes basée sur un thésaurus commun

Problèmes

- Ambiguïté des mots à différentes significations
- Difficulté de construction de bases de connaissances sans les contextes
- Synthèse des informations trouvées pour raffinement de l'aide (répondre efficacement aux questions)

Sujets de recherche (topics)

- Construction de sujets dérivés
- Définition de grammaires
- Détails avec des connecteurs et poids selon des facteurs de conformité
- Focalisation sur certains attributs
- Hiérarchie de concepts

Inférences

- Sortie de conclusions à partir de faits
- Intégration de nouveaux faits et de règles
- Moteur d'inférence décide sur les règles qui seront appliquées
- Arbres de décision consultés pour faire progresser la recherche d'informations
- Actions prédéterminées

Systemes à base de règles

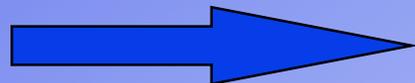
- Un système expert représente les connaissances en termes de **faits et règles**
- Consultation de la base pour évaluer les règles
- Deux techniques d'inférence : chaînage avant et chaînage arrière
- Lien sémantique de type is a :
 - Mars est une planète veut dire que mars hérite des propriétés planète
- Faits
 - l'entrepôt 41 est situé à Lennoxville
 - est-situé est un prédicat tandis que Entrepôt et Lennoxville sont des arguments
 - Columbia Picture est-une filiale de Coca-Cola
 - Le carbone est-un élément
 - un fait est utilisé pour une connaissance permanente ou temporaire

Règle: méthode de représentation de l'expertise

- SI consommation importante de sel
ALORS pression sanguine élevée
- Combinaison de règles avec les faits pour aboutir à une conclusion (nouveau fait)

- SI inflation est haute
- ALORS prix des maisons est haut

fait à vérifier
inflation est haute

 nouveau fait

- Prix des maisons est haut vrai si
inflation est haut vrai

- Ajout de faits : Assert

- Retrait de faits : Retract

2 prédicats


- Si la pluie est présent
Alors les nuages sont présents

Si A est vrai B est aussi vrai (noter que si A est faux B peut-être vrai)

- implication en avant : Si A alors B
- en arrière : B Si A

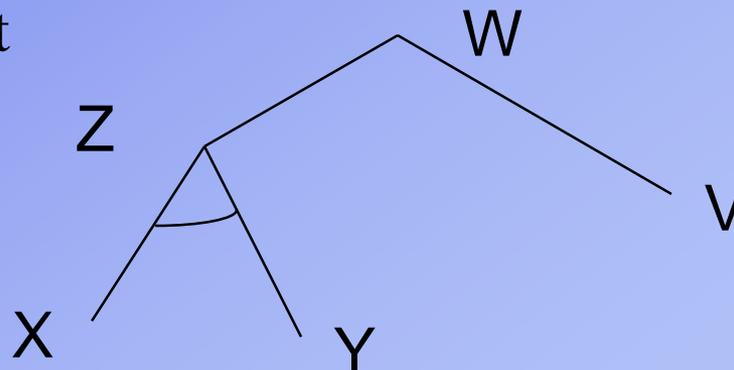
3 formes de règles

- Règles en **chainage avant**
SI pluie est présente alors nuages sont présents
- Règles en **chainage arrière**
nuages sont présents SI pluie est présente
- Règles inexactes
CF=25
pluie est présente SI nuages sont présents

Inférences

- Combinaisons de faits et de règles pour aboutir à de nouveaux faits: arbre de possibilités (arbre ET/OU)
racine : dernier but à prouver à travers
les noeuds : chemin de preuve (sous-arbre)
- **Chainage arrière:** départ de la racine et suivi des branches jusqu'aux feuilles jusqu'à vérification d'un fait
- **Chainage avant:** départ des feuilles vers la racine, suivi des branches vers le but

Si X And Y
ALORS Z;
Si Z ou V
ALORS W;



Chainage arrière, exemples

- But à prouver  focalisation sur les conclusions des règles
- vendeur A suggère argent
- vendeur A suggère or
- vendeur B suggère argent
- argent est surévalué
- or est sous-évalué
- **Règle 1** investisseur **investit dans** "X"
si vendeur A suggère "X"
et vendeur B suggère "X"
et non X est surévalué
- **Règle 2** investisseur **investit dans** "X"
si vendeur A suggère "X"
et X est sous-évalué

Faits

- But à prouver : investisseur investit dans or
- 1 - vérifier si ce but est atteint avec les faits et les règles présents
 - le fait n'est pas dans la base de fait
 - trouver une règle dont la conclusion est: investisseur investit dans or
- 1ère règle (X relié à or)

vendeur A <u>suggère</u> or	(1)	se trouve dans la base de faits
et vendeur B <u>suggère</u> or	(2)	ne vérifie pas base de faits ou la
et non or <u>est</u> surévalué		conclusion d'une règle
- 2ème règle: on doit prouver

vendeur A suggère or		2 clauses sont dans la base de faits
et or est sous-évalué		règle est prouvée

Chainage avant, exemple

- vendeur A suggère or
vendeur A suggère argent
vendeur A suggère banque nationale
vendeur B suggère argent
vendeur B suggère banque nationale
- Règle 1 si vendeur A suggère "X"
et vendeur B suggère "X"
alors investisseur investit dans "X"
- Règle 2 si "X" investit dans Y
alors X possède Y

- On prend chaque fait et on vérifie si les règles sont vraies
 - 1 - vendeur A suggère or (prochain but de la même règle)
 - vendeur B suggère or pas vérifié dans la base de faits
- vendeur A suggère argent
- vendeur B suggère argent
- vendeur A suggère argent OK, on rajoute la conclusion dans la base de faits
- investisseur investit dans argent
- 2ème règle : investisseur possède argent (plus de règles on retourne à BF)
- vendeur A suggère banque nationale, 2ème clause vérifiée vendeur B...
- investisseur investit dans Banque nationale rajouté dans la base de faits
- 2ème règle : investisseur possède Banque nationale rajouté dans la base de faits
- on cherche à impliquer ce nouveau fait
 - 1ère règle : pas de vérification pour X suggère Y
 - 2ème règle : pas de vérification, STOP.

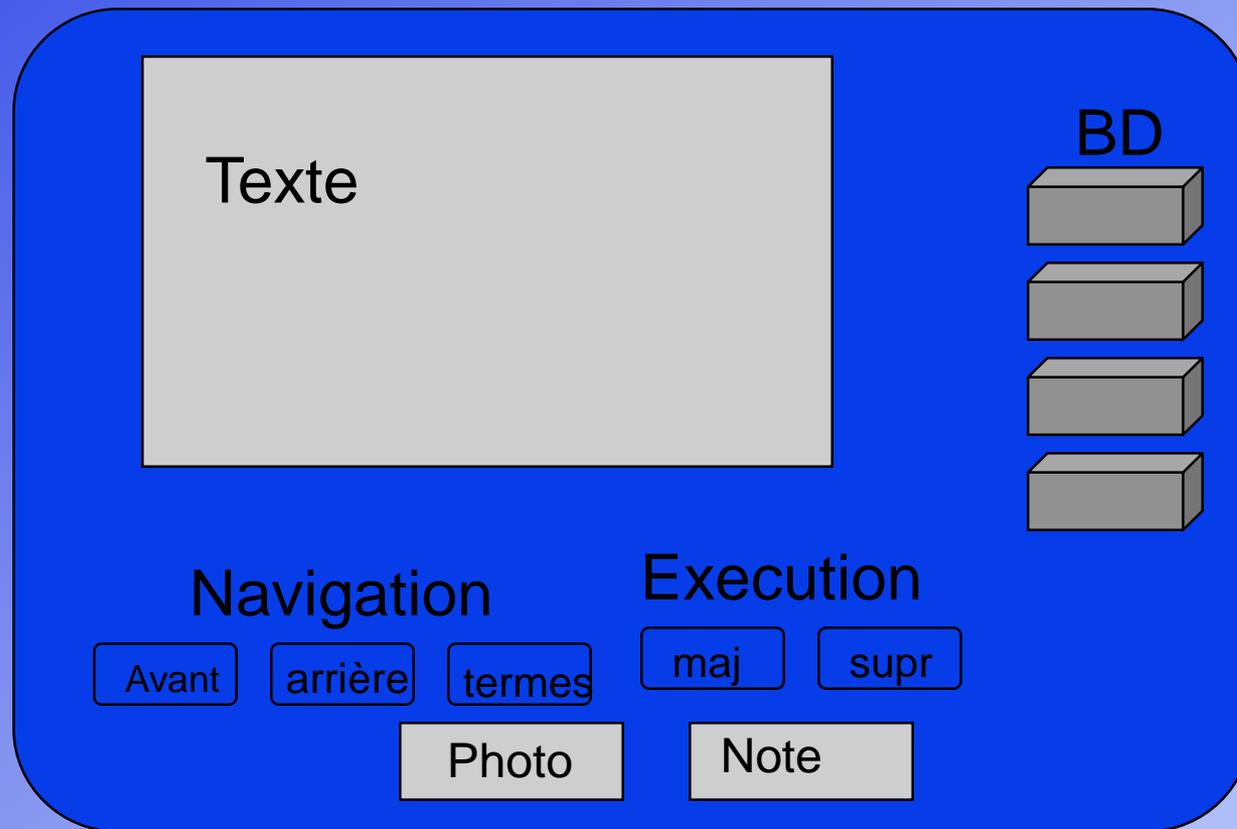
Problématique de l'acquisition de connaissances

- Les connaissances d'un expert sont subjectives et difficiles à formaliser.
- Les formalismes utilisés pour la représentation des connaissances ne permettent pas un bon niveau d'abstraction et restent liés à l'implantation.
- Comment organiser les connaissances en vue de leur traitement ?
- Transfert d'expertise => Modélisation des connaissances et de la méthode de résolution de problèmes

Interface intelligente

- Point critique des applications
- Mouvement et navigation
 - circuler entre objets
 - information sur l'objet
 - questionner
 - modifier
 - prendre notes objet courant
 - transfert des informations d'un objet vers une base de données

Exemple d'interface intelligente



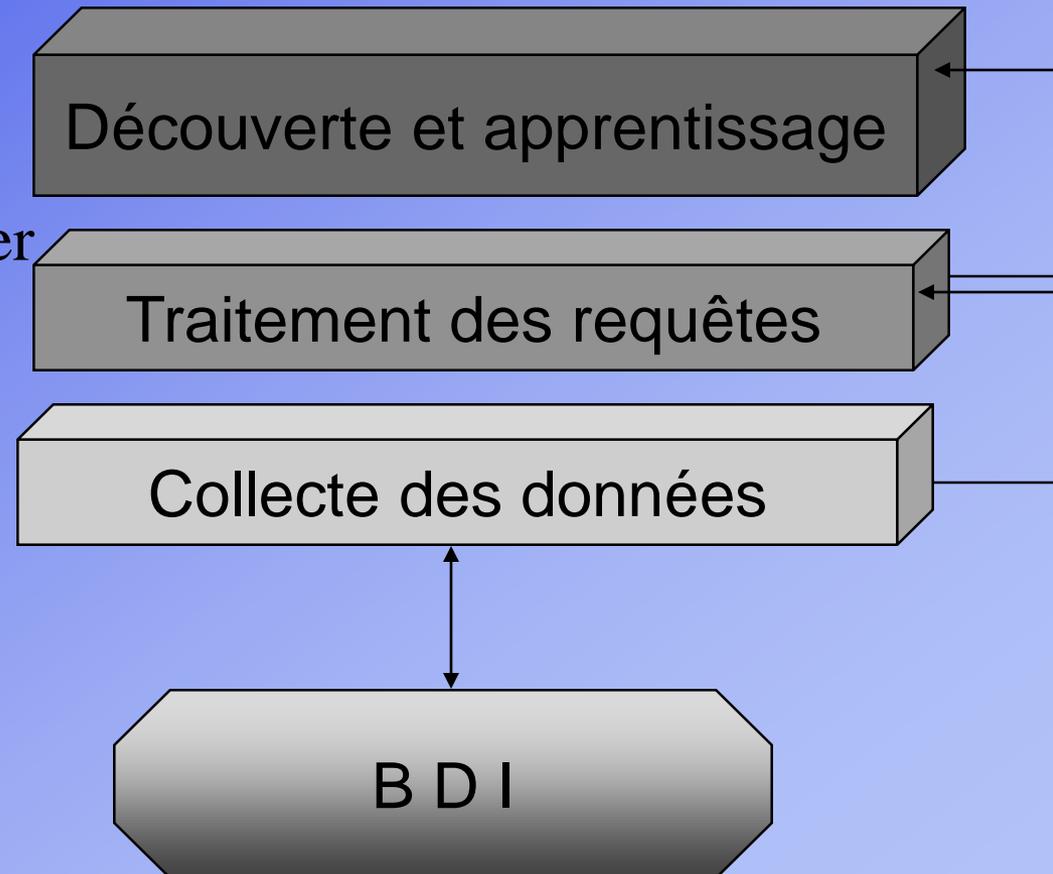
Outils de haut niveau

- **Outils de découverte**

- Apprentissage machine et découverte
- Utilisation de l'induction
- Permet à l'utilisateur de trouver automatiquement des informations cachées dans une grande base de données
- Découverte de relations entre les données par observation de modèles  programme d'analyse intelligent
- Ordinateurs sous-utilisés actuellement (30% du temps, 8h)
 - Volume important de données
 - Temps de traitement non utilisé
- En BD quelle question poser? En apprentissage le programme découvre les questions intéressantes à poser

Niveaux d'interaction

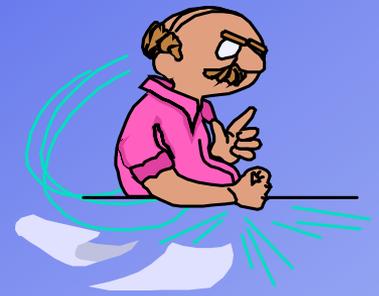
- Former des hypothèses
- Faire des requêtes
- Examiner résultats et modifier hypothèses
- Continuer le cycle jusqu'à apparition d'un modèle



Intégration de BD et de systèmes experts

- Deux types de couplage: faible et fort
- **couplage faible:** la base de données utilise le système expert pour analyser, déduire des données et réintroduire les données dans la base
- **couplage fort:** la base de données est une extension du système expert
- Opérateurs différents selon les approches

Méthodes statistiques



- méthodes pour analyser les données numériques (pas les connaissances)
- statistiques inférentielles reposent sur des modèles mathématiques de distribution des données
- analyse factorielle, analyse multivariante, analyse de séries chronologiques
- elles nécessitent une connaissance des statistiques et la présence d'un statisticien pour les analyser
- produisent les résultats sous forme d'équations or les règles sont plus lisibles que les équations
- elles évoluent vers l'IA

Apprentissage machine



- **Capacité d'un programme de découvrir et d'apprendre par lui-même**
- **ID3 (Quinlan, 1979, Michie, 1984)**
 - Certains bons résultats mais aussi une déviation dans les exemples peut produire de grands changements dans les arbres de décision
- **Extension avec AQ, INDUCE, CLUSTER (Michalski et al, 1983)**
- **RX et Radix (Blum, 1982-86) sont combinées avec des méthodes statistiques**
- **ENTAIL (Gaines et Shaw, 1996) est faite pour traiter des données inexactes**

- **ID3 : génère un arbre de décision**
- **AQ généralise des règles (plus puissant que ID3)**
- **INDUCE génère des règles pour des descriptions structurelles**
- **CLUSTER regroupe les données et découvre des structures**
- **RX, RADIX traite statistiquement les données temporelles**
- **ENTAIL génère des règles inexactes pour données inexactes**
- **Classification de Holland fournit des modèles**
- **IXL combine statistiques et apprentissage machine**
- **De nombreuses applications ont montré l'efficacité de IXL
industrie de fabrication, crime, médecine,....**

Intégrité de données et contrôle de qualité

- volume de données augmente et le nombre d'erreurs aussi
- nécessité d'avoir des outils qui détectent les erreurs
- contrôle d'intégrité au niveau de la définition de BD
- possibilité d'exprimer les contraintes en termes de règles ce qui donne plus de flexibilité
- outils statistiques pour mesurer la qualité des données

Outils de gestion hypermedia

- ensemble d'informations disponibles selon les intérêts des usagers. Doivent permettre d'effectuer diverses tâches
- représentation de connaissances
 - textes et graphiques
 - concepts
 - structures d'organisation
 - relations entre concepts
- notion de contexte

Ingénierie des connaissances/BDI

- **ingénierie des connaissances**

- o identifier connaissances
 - o expliciter connaissances
 - o structure de la connaissance
 - o représenter la connaissance
 - o Utiliser la connaissance
 - o valider, raffiner, maintenir connaissance
- 
- ```
graph TD; A[identifier connaissances] --> B[expliciter connaissances]; B --> C[structure de la connaissance]; C --> D[représenter la connaissance]; D --> E[Utiliser la connaissance]; E --> F[valider, raffiner, maintenir connaissance];
```

## **ingénierie des BDI**

- o identifier connaissances, données
  - o lien avec BDI
  - o construire interface hypermedia
  - o utiliser la BDI
  - o valider, raffiner, maintenir BDI
- 
- ```
graph TD; A[identifier connaissances, données] --> B[lien avec BDI]; B --> C[construire interface hypermedia]; C --> D[utiliser la BDI]; D --> E[valider, raffiner, maintenir BDI];
```

Outils d'aide à la décision

- les tableurs sont des outils qui permettent de répondre à une variété de questions du type: what if ?
- assurent une visibilité immédiate, une familiarité et une facilité de manipulation
- utilité de disposer aussi d'outils de transformation de formats pour passer d'un système à un autre (DB2, SQL,...) et notamment pour communiquer avec SQL
- utilité d'associer un outil de parcours (browser) avec une prise de notes et une recherche dans la BD + système d'inférence qui détermine le prochain noeud intéressant pour l'usager

- nécessité de connaître le profil de l'utilisateur (intérêts, objets de la requête)
- souvent recherche de l'information avec une préoccupation à l'esprit  prise de note utile
- stockage de notes comme des objets, indexage, édition, regroupement de notes par sujet
- annotation de photos (série de champs retrouvés par l'utilisateur) ou lien avec des notes existantes
- organisation hiérarchique des recherches d'information (détails progressifs fournis)

Systemes conseillers

- Implique l'intégration de systèmes experts et la conduite de dialogue avec l'utilisateur
- BD est considérée comme une extension de la base de connaissances
- différents niveaux de conseils
 - conseils sur demande de l'utilisateur: What to do next ?
 - conseil automatique: conseiller recommande car déduit des informations qui ne sont pas directement accessibles à l'utilisateur
 - conseil stratégique: après analyse des requêtes ou actions de l'utilisateur le système recommande les informations ou actions pertinentes

FIN