

RESUMO

Técnicas de inteligência artificial, aplicadas a dados de transações de empréstimo do acervo de uma biblioteca, podem gerar recomendações de itens relevantes para usuários e pesquisadores. O sistema sugerido neste relatório é baseado em procedimentos de consolidação e conexão de registros em um *data mart*, associados às técnicas de Análise de Cestas, Análise de Agrupamentos e Análise de Redes, numa estratégia de cooperação indireta.

No modelo aqui proposto, os itens da biblioteca são consolidados em Assuntos Significativos e Grupos Temáticos. Os usuários, por outro lado, são separados por Grupos Temáticos e segmentados segundo seu perfil de leitura.

Por meio de consolidação são criadas listas para cada Subgrupo Especializado obtido, contendo recomendações específicas, temáticas e gerais. Sugere-se a utilização de transações virtuais para aperfeiçoar as recomendações.

Além de servir a bibliotecas, o modelo proposto tem aplicabilidade direta em livrarias virtuais e pode ser adaptado para praticamente qualquer tipo de empreendimento em *e-business*.

PALAVRAS-CHAVE

Análise de Agrupamentos; Análise de Cestas; Análise de Redes; Bancos de Dados; Bibliotecas; Cooperação Indireta; Data Mart; Data Warehouse; Gestão de Conhecimento; Estatística; Inteligência Artificial; Livrarias Virtuais; Recomendações de Leitura; Marketing; Transações Virtuais.

ABSTRACT

Artificial Intelligence techniques can be applied to library circulation transactions to generate suggestions of items relevant to patrons and researchers. The Recommendation System presented in this report is based on consolidation and linkage of transaction records stored in a Circulation Data Mart, and on Basket Analysis, Cluster Analysis and Link Analysis techniques. An strategy of indirect cooperation has been adopted.

In the proposed model, library items are consolidated in Significant Subjects and Theme Groups. Patrons, on the other hand, are clustered by Theme Group and segmented by their reading profile.

Through consolidation, recommendation lists are generated for each Specialized Subgroup. Lists contain specialized, thematic and general suggestions. Virtual transactions are used to improve recommendations.

The Recommendation System developed in this project can be used not only in libraries but also in virtual bookshops; it can be easily adapted to practically any kind of e-business enterprise.

KEY WORDS

Artificial Intelligence; Basket Analysis; Cluster Analysis; Databases; Data Mart; Data Warehouse; Indirect Cooperation; Knowledge Management; Libraries; Link Analysis; Marketing; Reading Suggestions; Statistics; Virtual Bookshops; Virtual Transactions.

SUMÁRIO

I.	Retomada da relevância.....	7
1.	Do acervo ao usuário	8
2.	Informação personalizada: “Temos sugestões para você”	9
3.	Exemplo virtual	10
II.	Colaboração indireta.....	12
1.	Cooperação indireta e carga de trabalho	12
2.	Trabalhos recentes sobre cooperação indireta	13
3.	Colaboração indireta entre usuários de uma biblioteca.....	14
3.1.	Relevância geral <i>versus</i> relevância especializada.....	15
3.2.	Perfil dos usuários	15
3.3.	Grupos temáticos	16
III.	Biblioteca Karl A. Boedecker	17
1.	Acervo.....	17
2.	Usuários	18
3.	Transações.....	23
4.	<i>Data Mart</i> de Circulação	26
5.	Proposta do sistema de recomendações.....	28
IV.	Sistema de recomendações.....	28
1.	Definição dos assuntos significativos	29

2.	Identificação de grupos temáticos	33
2.1.	Análise de cestas com Model One	33
2.2.	Validação da análise do Model One.....	38
2.3.	Algumas considerações sobre os grupos temáticos	42
3.	Criação dos subgrupos especializados de pesquisadores para cada grupo temático.....	46
3.1.	Separação dos usuários com transações em um GT de interesse	46
3.2.	Agrupamento dos usuários conforme similaridade no perfil de assuntos significativos	46
4.	Criação de listas-base para o SGE	51
5.	Criação de listas temáticas	53
6.	Criação de uma lista geral.....	54
7.	Criação da lista personalizada.....	55
7.1.	Seleção de itens da lista especializada	56
8.	Exemplos dos resultados obtidos	56
9.	Transações virtuais	58
10.	Próximos passos	59
V.	Oportunidades de pesquisa	59
VI.	Conclusões finais.....	60
VII.	Bibliografia	61

ÍNDICE DE FIGURAS

Figura 1 – Protótipo de página de recomendações de títulos ao leitor.....	8
Figura 2 – “We think you’ll like these items”	11
Figura 3 – Transações diárias	23
Figura 4 – Transações por dia da semana.....	24
Figura 5 – Estrutura do <i>Data Mart</i> de circulação	27
Figura 6 – Análise de cesta de assuntos significativos (Model One/ArcView).....	34
Figura 7 – Algumas considerações sobre os resultados de análise de cesta (Model One/ArcView).....	38
Figura 8 – Análise de cesta de assuntos significativos (Minitab/Arcview).....	40
Figura 9 – Perfil de especialização dos grupos temáticos focados e genéricos.	44
Figura 10 – Perfil de especialização de grupos temáticos com distribuição bi-modal.	45
Figura 11 – Dendrograma dos usuários com transações no GT de marketing.....	47
Figura 12 – Assunto central <i>versus</i> assunto marginal.....	48
Figura 13 – Grupos muito especializados.	49
Figura 14 – Grupos pouco especializados.	50
Figura 15 – Grupos muito diversificados.....	51
Figura 16 – Sugestões personalizadas.....	57

ÍNDICE DE TABELAS

Tabela 1 – Exemplo de perfil de leitura de alguns usuários	16
Tabela 2 – Usuários registrados na Biblioteca Karl A. Boedecker, por categoria de vinculação	19
Tabela 3 – Número de usuários, por categoria com empréstimos no período estudado, em ordem decrescente de média de itens emprestados da categoria.....	21
Tabela 4 – Média de itens e assuntos tomados em empréstimo por alunos de pós-graduação, professores e funcionários, em ordem decrescente de número de grupos temáticos.....	22
Tabela 5 – Distribuição das transações de empréstimo pelos grupos de usuários conforme o dia da semana.....	25
Tabela 6 – Transações de empréstimo e devolução de itens	25
Tabela 7 – Exemplo de consolidação de “Códigos CDU” em “Assuntos Significativos”.....	31
Tabela 8 – Distribuição das transações por assuntos significativos	32
Tabela 9 – Grupos temáticos e correspondentes assuntos significativos, segundo Model One.....	36
Tabela 10 – Grupos temáticos e correspondentes assuntos significativos segundo análise de agrupamentos.....	41
Tabela 11 – Número de transações por grupo temático	43
Tabela 12 – Lista-base do SGE 10 do GT 1- Marketing.....	52
Tabela 13 – Lista de recomendações temáticas: SGE 1 – Marketing.....	54
Tabela 14 – Lista de recomendações gerais.....	55
Tabela 15 – Sugestões especializadas personalizadas para o usuário 410.....	56

PERFIL DE USUÁRIOS DA BIBLIOTECA KARL A. BOEDECKER*

Francisco Aranha¹

I. RETOMADA DA RELEVÂNCIA

Com a perda do monopólio sobre a informação, as bibliotecas tradicionais devem encontrar novas maneiras de agregar valor aos seus serviços para continuarem competitivas frente às diversificadas e sofisticadas fontes de informação hoje disponíveis aos usuários, particularmente aquelas que podem ser acessadas via Internet.

Felizmente, com a informatização dos catálogos e dos processos de gestão do acervo, a aplicação de novas tecnologias de gestão do conhecimento tornou-se disponível para as bibliotecas, a um custo relativamente baixo.

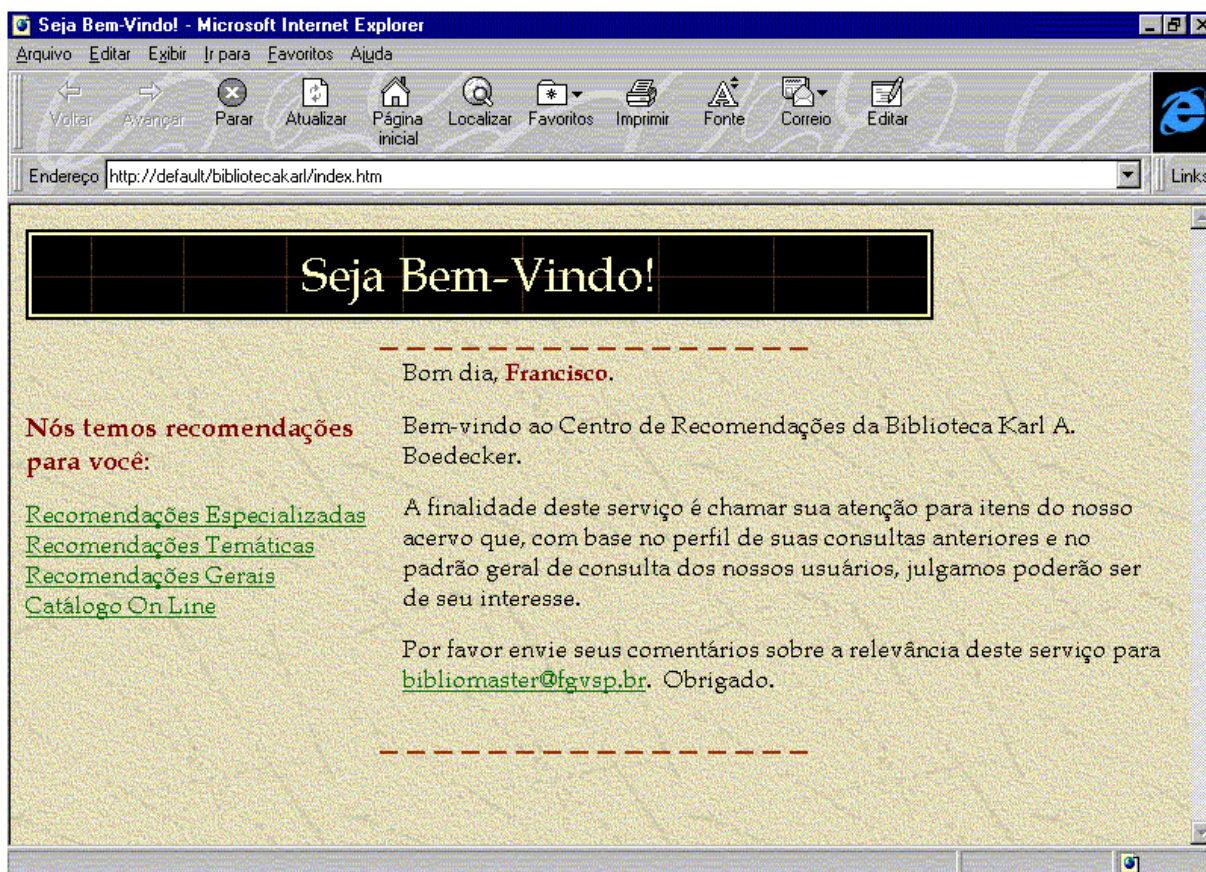
A incorporação destas tecnologia pode transformar a ameaça da Internet em uma oportunidade de alavancagem da relevância da biblioteca para a sua comunidade de usuários. Neste trabalho, propomos que a potencialização do uso da informação armazenada no acervo seja obtida pela facilidade de **acesso** (WURMAN, 1991) do usuário a itens que lhe sejam relevantes. Na Figura 1, apresentamos um protótipo da *Home Page* da Central de Recomendações proposta neste projeto.

* O NPP agradece aos alunos que participaram da pesquisa que originou o presente relatório como auxiliar de pesquisas, Vera Lúcia de Góes, e como monitor de pesquisas, Érico Resende Santos.

¹ Colaboraram: Érico Resende Santos (autor de projeto de pesquisa correlato, cujos resultados serviram de suporte à realização deste trabalho) e Vera Lúcia de Góes (auxiliar de pesquisa). O autor gostaria também de registrar e agradecer o apoio financeiro do Núcleo de Pesquisa e Publicações da EAESP/FGV, que custeou este projeto; e o suporte da empresa Informix do Brasil, que forneceu a infra-estrutura tecnológica necessária à implementação do *Data Mart* de Circulação.

Figura 1

Protótipo de Página de Recomendações de Títulos ao Leitor



1. DO ACERVO AO USUÁRIO

Uma atitude apenas reativa aos marcados avanços da tecnologia na área de Gestão de Conhecimento pode fornecer uma sobrevida ao modelo atual de atendimento de usuários de bibliotecas. No entanto, somente uma ativa reformulação de seu papel e objetivos poderá garantir a continuidade a longo prazo destas instituições (CARSON

et al., 1997), uma vez que outras fontes de informação competirão pela preferência do usuário.

Uma das mais profundas e indispensáveis mudanças do foco das bibliotecas será o deslocamento da ênfase na organização de um acervo restrito em número de itens e tipos de mídia, para a ênfase na gestão da abundância excessiva de informações de todas as fontes e formatos. Na administração de grandes volumes de dados, destacar-se-ão as habilidades de filtro, seleção e empacotamento personalizado dos itens relevantes ao usuário.

Para viabilizar esta nova abordagem, as bibliotecas deverão, cada vez mais, ser capazes de **concentrar atenção no usuário e seu comportamento**, em oposição a **concentrar-se exclusivamente no acervo**. Os profissionais da informação deverão deixar de ser porteiros (“*gatekeepers*”), meros organizadores e controladores do acesso às estantes, para serem portais (“*gateways*”), isto é, mapeadores e auditores de fontes internas e externas de informações relevantes ao usuário (Stephen Abram, segundo MILLER, 1998).

2. INFORMAÇÃO PERSONALIZADA: “TEMOS SUGESTÕES PARA VOCÊ”

“A segmentação avançou tanto nesta década de 90, que nenhuma empresa [ou instituição] pode ignorá-la” (KOTLER, 1997, p. 136).

Isto inclui as bibliotecas, que serão compelidas a implantar uma estratégia de relacionamento com usuários baseada nos princípios do marketing um-a-um. Estes se traduzem na “disposição e possibilidade da organização em mudar de comportamento com relação a usuários individuais, com base no que o usuário diz e no que se sabe sobre ele” (PEPPERS e ROGERS, 1999, p. 151) .

3. EXEMPLO VIRTUAL

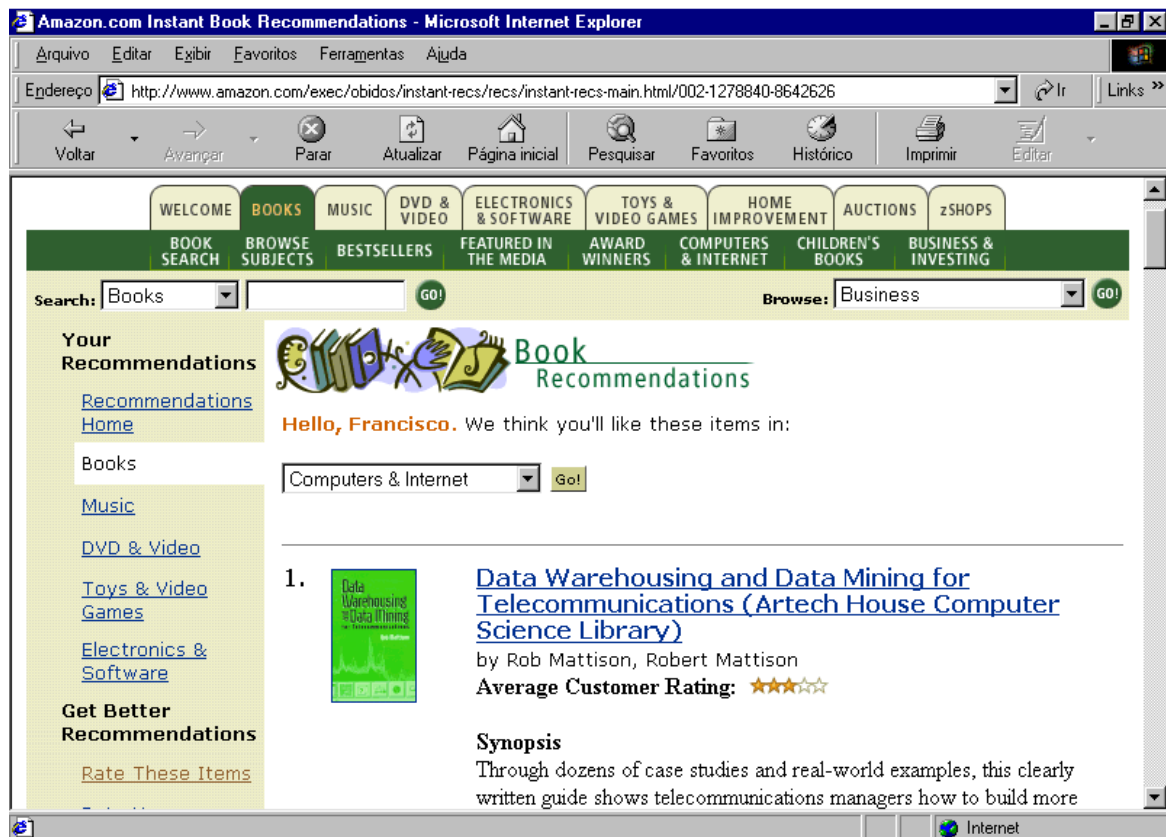
As mesmas forças de mudança social e tecnológica que estão agindo sobre as bibliotecas atuam também sobre empreendimentos correlatos, as livrarias.

Fortemente expostas à competição e mais sensíveis às mudanças de hábito dos consumidores do que as bibliotecas, as livrarias têm sido impelidas a se transformar mais rapidamente, inclusive implementando estratégias de comércio eletrônico. Segundo o criador da Amazon.com, Jeff Bezos, o principal motivo pelo qual clientes voltam a uma livraria virtual é **a ajuda ativa** que o *site* oferece na localização do item procurado, mesmo quando a existência do item é desconhecida para o próprio usuário, e sua necessidade só se torna clara para ele quando recebe a sugestão e reconhece nela inegável relevância.

A localização destes itens apenas vagamente especificados pode ser feita porque o *site* conhece tão bem o perfil e os hábitos de consumo do cliente, que identifica itens relevantes e os leva sua à sua atenção (RAMO, 1999).

Figura 2

“We Think You’ll Like These Items”



Fonte: <http://www.amazon.com>

Para evitar uma drástica redução no número de consultas, em breve também as bibliotecas serão compelidas a agir pró-ativamente. Felizmente, isto não será difícil: em decorrência de sua capacidade de observação do comportamento dos usuários, as bibliotecas poderão, com facilidade, implementar estratégias de colaboração indireta, como as sugeridas neste projeto de pesquisa.

II. COLABORAÇÃO INDIRETA

David PAYTON (1998), em artigo sobre a identificação de similaridade entre pesquisadores por meio da comparação das páginas que consultam na Internet, define colaboração indireta como o procedimento:

“que transforma as atividades coletivas de um grupo em fonte de informação personalizada para seus membros, permitindo a cada um beneficiar-se de maneira particular do conhecimento implícito no processo de coleta e consolidação de informação dos demais membros do grupo” (PAYTON, 1998, p. 1).

1. COOPERAÇÃO INDIRETA E CARGA DE TRABALHO

Os procedimentos de cooperação indireta tendem a ser bem sucedidos quando a captura das informações necessárias à sua implementação é realizada por um mecanismo automatizado, que registra as atividades comuns e cotidianas dos membros do grupo, sem lhes atribuir uma carga de trabalho adicional. Caso contrário, o ônus de preencher relatórios, documentar comportamentos ou relatar achados de pesquisa desencoraja a participação no sistema (SCHWARTZ e WOOD, 1993).

Se o esforço para participar é alto, o procedimento aproxima-se dos sistemas de cooperação direta, em que os atores tentam ajudar-se mutuamente por meio de ações intencionais ou comunicações explícitas, como, por exemplo, participações em conferências ou troca de correspondência.

As iniciativas de cooperação podem variar amplamente nestes eixos de intencionalidade e carga de trabalho. Na travessia de uma floresta desconhecida, por exemplo, um viajante poderia beneficiar-se da cooperação direta de um guia.

Alternativamente, poderia utilizar a cooperação indireta, mas ainda trabalho-intensiva, de um sistema de sinalização por placas instaladas pela comunidade de viajantes, com cada membro se responsabilizando pela identificação de uma pequena região da floresta; neste caso, o volume de trabalho é menor do que o necessário para cada viajante ter seu próprio guia. Finalmente, em um sistema de cooperação indireta passivo, o viajante seria guiado pelos sinais de desgaste de trilhas e picadas demarcadas apenas pelo uso freqüente de outras pessoas que cruzaram a mesma região.

Instituições que realizam transações eletrônicas podem facilmente utilizá-las para gerar em seus bancos de conhecimento estas “trilhas delimitadas por desgaste”, passíveis de interpretação.

A explicitação do conhecimento sobre o comportamento grupal e a seleção da parcela deste conhecimento relevante para cada usuário em particular é uma das estratégias utilizadas pela Amazon.com em suas recomendações, e pode ser empregada igualmente por bibliotecas.

2. TRABALHOS RECENTES SOBRE COOPERAÇÃO INDIRETA

Várias iniciativas de implementação de sistemas de cooperação indireta foram documentadas recentemente. PAYTON (1998) procurou facilitar o contato entre pessoas com interesses comuns explicitando seu padrão de navegação na Internet. KAUTZ, SELMAN e SHAH (1997a, 1997b) caracterizaram redes de pesquisadores que mantinham vínculos sociais, analisando a co-ocorrência de nomes em documentos públicos na Internet. SCHWARTZ e WOOD (1993) identificaram colaboradores potenciais pela análise do tráfico de emails em pontos selecionados da rede. SWANSON e SMALHEISER (1999) desenvolveram o software Arrowsmith

para identificar justaposições sugestivas de títulos de artigos na área de biomédicas, indicativas de relações pouco evidentes entre achados científicos.

Baseados nos mesmos princípios de cooperação indireta, mas caminhando em direção oposta, isto é, perseguindo o objetivo de **dificultar** a cooperação, há vários trabalhos na área de detecção de fraudes contra companhias seguradoras e de saúde (CABENA e OUTROS, 1998), combate à lavagem de dinheiro e ao crime organizado em geral (JENSEN, 1997; HANN, 1998).

3. COLABORAÇÃO INDIRETA ENTRE USUÁRIOS DE UMA BIBLIOTECA

Tipicamente, um pesquisador inicia seus trabalhos numa biblioteca identificando uns poucos itens claramente relevantes ao seu tema de pesquisa. A partir deles, segue as indicações explícitas dos autores, que citam suas fontes. Mapeando as citações, o pesquisador reconstrói uma teia de vínculos temáticos; computando referências cruzadas, estabelece níveis de relevância das obras citadas e orienta sua própria leitura (ECO, 1982).

A estratégia proposta neste projeto de pesquisa é rebater esta estratégia de cooperação direta, ativa e protagonizada pelo autor que cita, para uma estratégia **indireta, passiva e protagonizada pelo pesquisador que segue as citações**.

Admitindo-se que o pesquisador tenha um projeto de estudo coerente, quando examina dois trabalhos, estabelece um vínculo temático entre eles. Os trabalhos lidos por muitos pesquisadores são implicitamente os mais relevantes, da mesma maneira que são os trabalhos muito citados.

Assim, um sistema de cooperação indireta dos usuários de uma biblioteca deve disponibilizar informação sobre os itens mais consultados e sobre os itens

consultados em conjunto com outros itens, organizando esta informação por áreas temáticas relevantes a cada usuário em particular.

3.1. Relevância geral *versus* relevância especializada

A frequência de utilização de um item como medida de relevância será considerada em duas dimensões, geral e especializada:

- a relevância geral será indicada pela posição do item no ranking de utilização pelo conjunto de todos os usuários;
- a relevância especializada será indicada pela posição de um item no ranking de utilização apenas pelos membros de um Subgrupo Especializado (SGE) de pesquisadores.

Mesmo que um item seja muito utilizado em geral, ele pode ser pouco relevante para um grupo de pesquisadores com interesse bem delimitado. Por outro lado, um item pouco utilizado em geral, mas muito utilizado por um pequeno grupo de pesquisadores com interesses específicos pode ser altamente relevante para os membros deste grupo que ainda não o consultaram. De fato, a **diferença de posição** entre estes dois rankings, padronizada para refletir o tamanho relativo do ranking, é um indicador importante de relevância especializada.

3.2. Perfil dos usuários

Os usuários serão colocados em Subgrupos Especializados (SGE) com base em seu Perfil de Leitura – caracterizado pela distribuição de seus itens de consulta em Assuntos Significativos (veja item IV.1).

Tabela 1

Exemplo de Perfil de Leitura de Alguns Usuários

patron_key	a0	a1	a2	a3	a30	a31	a32	a33	a3301	a3303	a331	...	a6587	a6588	a659	a68	a7	a8	a9
5935	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4145	0	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0
4204	0	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0
4221	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4222	0	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0
4224	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4312	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4330	0	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4337	0	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0
4339	0	0	0	0	0	0	0	1	1	0	0	...	0	1	0	0	0	0	0
4439	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4510	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

Fonte: *Data Mart* de Circulação

Na Tabela 1 está representado um segmento de uma tabela, construída com base nos princípios de conexão (*linking*), em que se constrói uma rede de relacionamentos entre registros para a criação de um padrão de atividade (GOLDBERG e SENATOR, 1995).

No corpo desta tabela, o valor 0 indica que o usuário (linha) não retirou itens relativos a um assunto significativo (coluna). O valor 1 indica a existência de empréstimos de itens naquela área temática.

A seqüência de zeros e uns de um usuário caracteriza seu padrão temático de consultas.

3.3. Grupos temáticos

Grupos Temáticos são características dos itens da biblioteca. Sua definição procura refletir o campo de conhecimento abrangido pelo item. A identificação dos GTs

parte da Classificação Decimal Universal (IBICT, 1987; SILVA, 1994), mas é conduzida de forma a refletir a maneira como, na prática, os pesquisadores agrupam diversos assuntos em grandes áreas de pesquisa (veja item IV.2). Segundo esta abordagem, não é o bibliotecário quem define o Grupo Temático de um livro, mas os próprios pesquisadores, pelo uso que fazem do item.

Assim, por exemplo, se itens relativos à ciência política são utilizados por pesquisadores de Administração Pública, estes item são filiados ao GT de Administração Pública e não a um GT de Ciências Sociais, como aconteceria se fosse adotado integralmente o procedimento normativo indicado pela CDU.

Outra vantagem desta estratégia de seguir o uso dos pesquisadores é permitir mudanças na configuração dos GTs ao longo do tempo.

III. BIBLIOTECA KARL A. BOEDECKER

A Biblioteca Karl A. Boedecker foi criada em 1954 com o objetivo de fornecer apoio bibliográfico às atividades de ensino e pesquisa desenvolvidas pela comunidade acadêmica da Escola de Administração de Empresas de São Paulo (EAESP).

1. ACERVO

Seu acervo se compõe de cerca de 67 mil exemplares de 47 mil títulos de livros; e de 3,2 mil exemplares de 1,8 mil títulos de teses e dissertações (BIBLIOTECA KARL A. BOEDECKER, 1998).

Deste total de cerca de 70.000 itens, em 19/07/99, data de coleta dos dados utilizados neste projeto de pesquisa, 26.000 haviam sido incorporados ao VTLS,

sistema informatizado desenvolvido pela Virginia Tech University para consulta on-line de catálogos e automatização de controles de administração de acervos (DONNARD e BOREGAS, 1998; VTLS INC, 1999).

2. USUÁRIOS

Os cerca de 6.000 usuários da Biblioteca Karl Boedecker registrados no VTLS em 19/07/99 encontravam-se distribuídos pelas categorias de alunos, professores, funcionários e outros conforme Tabela 2, a seguir.

Tabela 2

Usuários Registrados na Biblioteca Karl A. Boedecker, por Categoria de Vinculação

Tipo de Usuário	Número de Usuários
Alunos	5,392
Graduação	1,652
Aluno de CGAE	1,233
Aluno de CGAP	419
Pós Graduação Lato Senso	2,973
Aluno de CEAG	1,741
Aluno de CEAHS (Proahsa)	60
Aluno de PECE	1,172
Pós Graduação Estrito Senso	527
Aluno de Doutorado	140
Aluno de Mestrado	264
Aluno de MBA	123
Outros Alunos	240
Aluno Doador	66
Ex-aluno	144
Aluno de Intercâmbio Internacional	30
Professor	304
Funcionário	276
Bibliotecas/Empresas/Outros	122
Total Geral	6,094

Fonte: *Data Mart* de Circulação²

Quando um usuário se desliga da biblioteca, tipicamente por concluir seu curso, o registro correspondente não é eliminado do sistema. Por este motivo, dos cerca de 6.000 usuários registrados, apenas cerca de 2.400 eram usuários ativos (BIBLIOTECA KARL A. BOEDECKER, 1998).

² Significado das siglas de uso corrente na Escola: CGAE – Curso de Graduação em Administração de Empresas; CGAP – Curso de Graduação em Administração Pública; CEAG – Curso de Especialização em Administração para Graduados; CEAHS – Curso de Especialização em Administração Hospitalar e de Saúde; PECE – Programa de Educação Continuada para Executivos.

Os leitores ativos podem freqüentar a biblioteca, consultar o acervo, tirar cópias de textos selecionados ou utilizar as salas de estudo em grupo; não necessariamente precisam retirar itens em empréstimo. Se o usuário não retirar um item, na estratégia atual de coleta de informações da biblioteca sua presença e atividades não ficam registradas no sistema. Assim, apenas as **transações de empréstimo** permitem associar o item ao usuário que o utilizou e, por este motivo, foram escolhidas para objeto deste estudo.

No período analisado (veja item 3, a seguir), 1.644 usuários ativos realizaram empréstimos (Tabela 3, a seguir). Constatou-se que, destes, os alunos de pós-graduação estrito senso, alunos do CEAHS, professores e funcionários tinham um padrão de freqüência de empréstimos significativamente distinto do padrão de alunos de cursos de especialização lato senso (exceto CEAHS) e de graduação. No período estudado, o primeiro grupo retirou, em média, aproximadamente o dobro do número de itens retirados pelo segundo grupo.

Tabela 3

Número de Usuários, por Categoria com Empréstimos no Período Estudado, em Ordem Decrescente de Média de Itens Emprestados da Categoria

Tipo de Usuário	Usuários	Média de Empréstimos no Período Analisado
Aluno de Doutorado	60	13.77
Aluno de MBA	39	13.54
Aluno de Mestrado	148	12.91
Aluno de CEAHS (Proahsa)	26	12.69
Professor	62	11.85
Funcionário	61	11.33
Aluno de CGAP	259	7.82
Ex-aluno	22	7.23
Aluno de Intercâmbio Internacional	13	7.15
Aluno de CGAE	621	6.63
Aluno doador	27	5.48
Bibliotecas/Empresas	25	4.88
Aluno de CEAG	260	4.08
Aluno de PECE	21	3.86
Total	1644	

Fonte: *Data Mart* de Circulação

Além da diferença com relação ao número de itens, constatou-se que os alunos de graduação apresentam um perfil de consulta envolvendo um elevado número de Grupos Temáticos (assuntos) distintos. Este perfil, no entanto, é razoavelmente homogêneo entre alunos de mesmo semestre, o que sugere ser a bibliografia consultada aquela indicada nos cursos regulares.

Os alunos de pós-graduação e do CEAHS, professores e funcionários, ao contrário, apresentaram um perfil temático razoavelmente restrito, caracterizado na Tabela 4, a seguir.

Tabela 4**Média de Itens e Assuntos Tomados em Empréstimo por Alunos de Pós-Graduação, Professores e Funcionários, em Ordem Decrescente de Número de Grupos Temáticos**

Tip de Usuário	No. de Usuários	Média de Itens	Média de Assuntos
Aluno de CEAHS	27	12.69	4.15
Aluno de Mestrado	150	12.91	3.44
Aluno de MBA	41	13.54	3.24
Funcionário	63	11.33	3.17
Professor	67	11.85	2.88
Aluno de Doutorado	62	13.77	2.79
Total	410		

Fonte: *Data Mart* de Circulação

Ainda com referência à Tabela 4, note-se que a área temática de consulta fica progressivamente mais focada conforme o aluno mais se especializa, passando pelos níveis de pós-graduação lato senso, mestrado, MBA e doutorado. Segundo este indicador, o grau de especialização dos professores é análogo ao dos alunos de doutorado, o que é razoável. Apesar de teoricamente coerentes, estes resultados são estatisticamente pouco significativos (valor-p de 19%) devido à grande variabilidade interna no número de assuntos consultados pelos grupos.

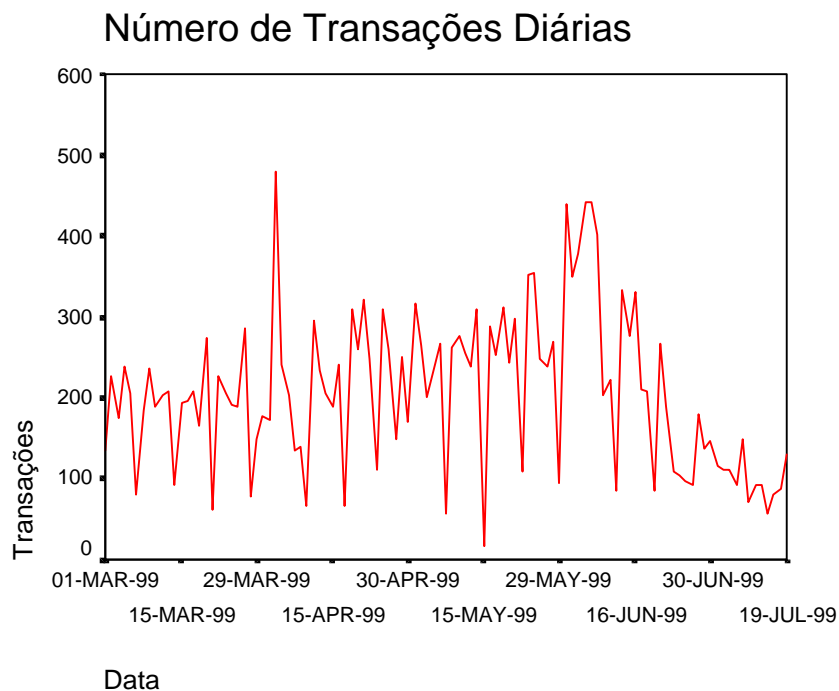
Tendo em vista a maior consistência temática, a maior intensidade de uso da biblioteca, e o maior benefício potencial de um Sistema de Recomendações para o grupo de alunos de pós-graduação, professores e funcionários, estes foram isolados para as análises de perfil e criação de listas de sugestão. Assim, neste projeto, serão preparadas listas de recomendações apenas para os 410 usuários considerados na Tabela 4.

3. TRANSAÇÕES

Para a realização deste projeto de pesquisa, foram capturadas informações sobre cerca de 22.500 transações de circulação de itens, ocorridas no período de 01/03/99 (data de implantação do sub-sistema de estatísticas do VTLS, que passou a registrar as transações) e 19/07/99 (data de coleta dos dados). A evolução do número de transações diárias está representada na Figura 3 em que podemos notar picos de movimento às segundas-feiras e vales aos sábados. Nota-se também um progressivo “aquecimento” do movimento durante o semestre, até antes dos exames; e um “desaquecimento”, conforme se avança no período de férias.

Figura 3

Transações Diárias

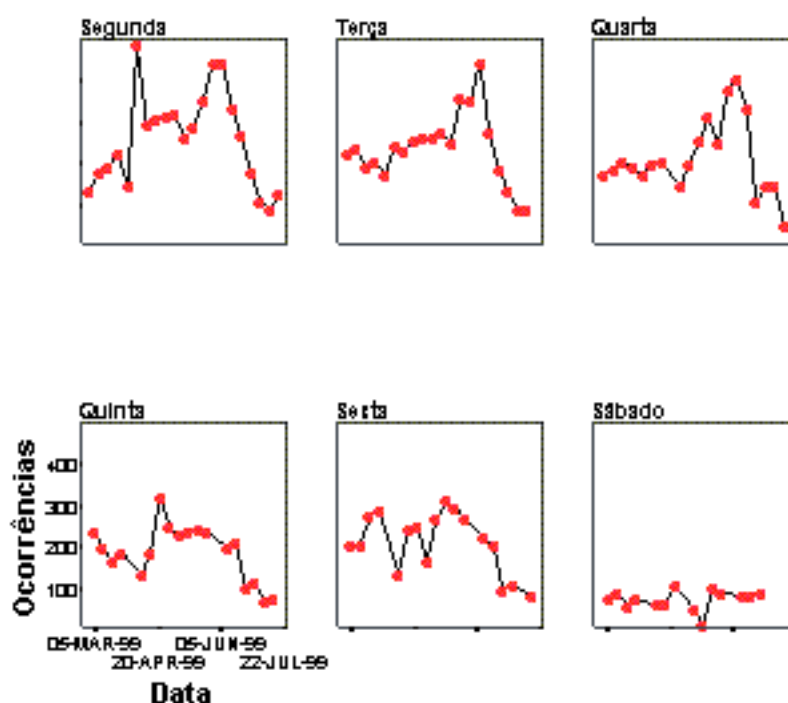


Fonte: *Data Mart* de Circulação

O aquecimento é mais notado no movimento de início de semana (segundas, terças e quartas-feiras), enquanto o movimento em finais de semana (quintas e sextas-feiras) tem uma frequência mais estável. O movimento aos sábados é praticamente estável ao longo do semestre.

Figura 4

Transações por Dia da Semana



Fonte: *Data Mart* de Circulação

Investigações adicionais indicaram que a diferença de sazonalidade no volume de transações conforme o dia da semana reflete uma diferente composição de grupos de usuários. Comparando-se os casos extremos (Tabela 5), observa-se que às segundas-feiras predominam os alunos de graduação enquanto aos sábados predominam os

alunos de mestrado e MBA; aos sábados também aumenta significativamente a frequência de usuários externos à Escola.

Tabela 5

Distribuição das Transações de Empréstimo pelos Grupos de Usuários conforme o Dia da Semana

Segunda Feira			Sábado		
Tipo de Usuário	Transações	%	Tipo de Usuário	Transações	% Sáb/Seg
MBA	157	3%	MBA	164	15%
Aluno doador	55	1%	Aluno doador	43	4%
Mestrado	710	13%	Mestrado	237	22%
Bibliotecas/empresas externas	42	1%	Bibliotecas/empresas externas	11	1%
Ex-aluno	96	2%	Ex-aluno	23	2%
CEAHS (Proahsa)	179	3%	CEAHS (Proahsa)	41	4%
CEAG/CEAG-Turismo/CEAG-Esporte	517	9%	CEAG/CEAG-Turismo/CEAG-Esporte	108	10%
Doutorado	376	7%	Doutorado	63	6%
CGAE	1720	31%	CGAE	232	21%
CGAP	958	17%	CGAP	129	12%
PIM Programa Intercâmbio Inter	33	1%	PIM Programa Intercâmbio Inter	3	0%
Professor	293	5%	Professor	23	2%
Funcionário	335	6%	Funcionário	20	2%
PECE	28	1%	PECE	0	0%
Total	5499			1097	

Fonte: *Data Mart* de Circulação

Finalmente, note-se que as transações distribuem-se entre categorias de empréstimo e devolução, conforme Tabela 6 a seguir.

Tabela 6

Transações de Empréstimo e Devolução de Itens

Transações de Circulação	Número de Transações
Devolução	9,682
Empréstimo	9,361
Renovação	3,465
Total	22,508

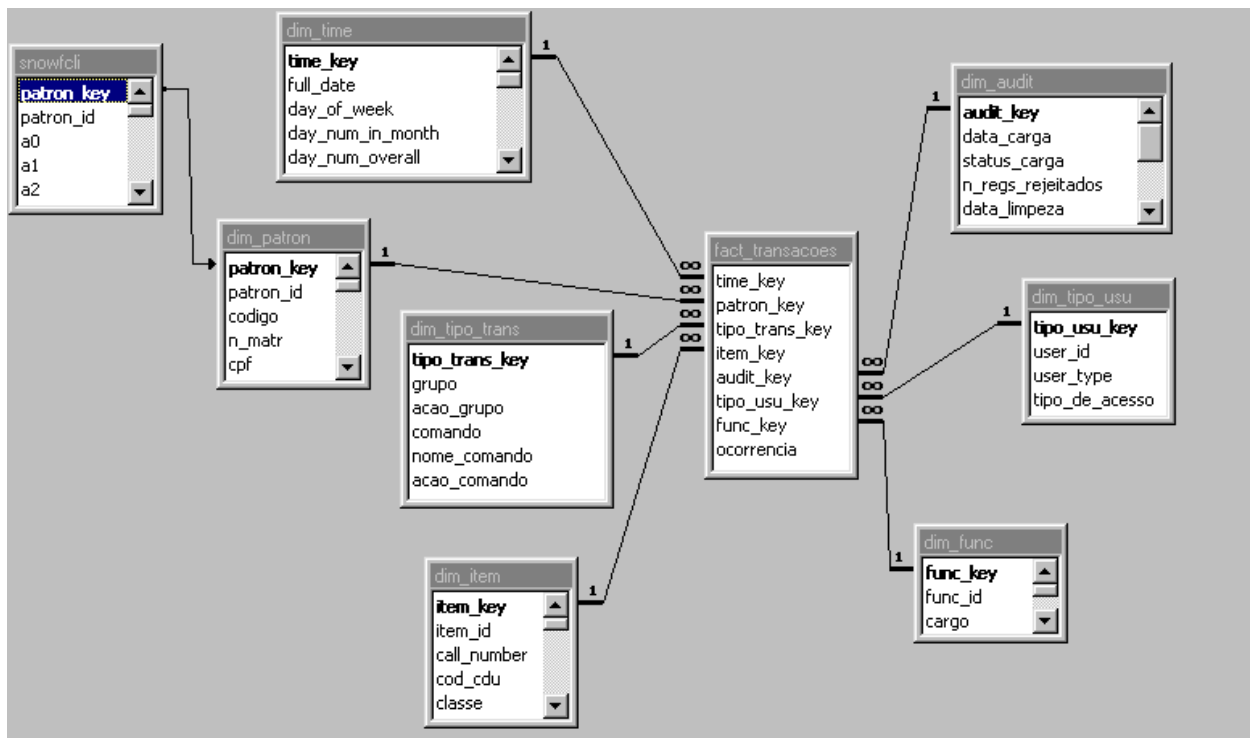
Fonte: *Data Mart* de Circulação

Para se evitarem problemas de dupla contagem, na elaboração de perfis de leitor foram analisadas apenas as transações de devolução final do item.

4. *DATA MART* DE CIRCULAÇÃO

No período de Agosto a Novembro de 1999, um projeto de pesquisa conexo a este, conduzido por SANTOS (1999), permitiu que os dados coletados no sistema VTLS fossem consolidados e organizados em um *data mart* piloto (KIMBALL, 1998; BERSON e SMITH, 1997). Nas análises apresentadas neste relatório foram utilizados os dados disponibilizados no *Data Mart* de Circulação, cuja estrutura está detalhada na Figura 5.

Figura 5

Estrutura do *Data Mart* de Circulação

Fonte: Adaptado de SANTOS, 1999

No centro da Figura 5 temos uma tabela-fato, contendo as transações de empréstimo; conectada à tabela-fato, temos, em sentido anti-horário a partir do canto inferior direito, as dimensões de funcionários, tipo de usuários, auditoria, tempo, usuários, tipo de transações e itens. Anexo à dimensão de usuários, temos um “*snowflake*” contendo seu perfil de consultas temáticas.

5. PROPOSTA DO SISTEMA DE RECOMENDAÇÕES

O Sistema de Recomendações (SR) da Biblioteca Karl Boedecker deve oferecer a cada um dos 410 usuários selecionados no item 2 uma lista personalizada de recomendações de títulos potencialmente de seu interesse. Esta lista será composta de:

- Itens de Interesse Especializado,
- Itens de Interesse Temático e
- Itens de Interesse Geral.

As diferenças entre estes seguimentos da lista, e a forma de prepará-los idealizada e testada neste projeto, será descrita no item IV, a seguir.

Uma vez criadas, as listas serão incorporadas ao *Data Mart*, onde permanecerão disponíveis para consulta apenas pelo próprio usuário, devendo ser atualizadas periodicamente por meio de procedimentos armazenados.

IV. SISTEMA DE RECOMENDAÇÕES

A implementação das listas personalizadas de recomendações descritas no item III.5 envolve as seguintes etapas:

- definição dos assuntos significativos;
- identificação de Grupos Temáticos (GT);
- criação de Subgrupos Especializados (SGE) de usuários para cada um dos GTs; e
- criação das listas individualizadas.

O usuário disporá de uma lista personalizada para cada GT de que participa.

1. DEFINIÇÃO DOS ASSUNTOS SIGNIFICATIVOS

Diante da necessidade de organizar seu acervo de uma forma inteligível e sistemática, as bibliotecas desenvolveram sistemas de classificação. A Biblioteca Karl Boedecker utiliza a CDU – Classificação Decimal Universal.

Derivada da CDD – Classificação Decimal de Dewey, a CDU é um esquema para a classificação de informações sobre todos os assuntos, quaisquer que sejam os tipos de documentos em que se achem registradas. Estabelece uma hierarquia entre os ramos do saber, de forma a permitir relacionar todos campos do conhecimento humano a um plano lógico, progressivamente mais detalhado (IBICT, 1987).

O processo de classificação, no entanto, revela apenas uma parte dos recursos totais existentes em um acervo; numerosas são as razões que a tornam incompleta, inflexível e pouco natural. Em consequência, e com a disponibilização progressiva de recursos de informática, a classificação tem perdido importância para o processo de indexação, que permite relacionar um item a vários temas e assuntos para os quais é relevante (DYER, BROWN e GOLDSTEIN, 1970).

De qualquer forma, justamente a rigidez e a hierarquia do CDU foi de interesse para o nosso projeto, permitindo-nos aplicar o procedimento de **consolidação**, típico de técnicas de Data Mining.

A consolidação permite fornecer um chave única a múltiplas versões de uma mesma entidade, presente no banco de dados de forma excessivamente detalhada ou ambígua (GOLDBERG e SENATOR, 1995).

Um exemplo de consolidação é a redução de registros de um mesmo usuário, multiplicados em consequência de variações na grafia de seu nome, a um registro único, que corresponde à entidade do mundo real “indivíduo”. Outro exemplo bastante típico é encontrado em técnicas de Análises de Cestas de Produtos:

diferentes mercadorias (análogas, no nosso caso, a itens de informação como livros ou teses) são associadas a uma taxonomia, isto é, um esquema de classificação hierarquizado, segundo a qual se define o nível de detalhe conveniente à análise. Geralmente deve-se fazer um *trade-off* entre detalhamento e complexidade analítica. Conforme se “sobe no esquema taxonômico”, reduz-se o número de itens a serem considerados (BERRY e LINOFF, 1997). Assim, por exemplo, em vez de se estudarem os produtos Toddy, Nescau e Muky, pode-se considerar a todos como realizações de uma mesma entidade, a classe dos achocolatados.

No caso da Biblioteca Karl Boedecker, o acervo cadastrado, de cerca de 26.000 itens, está distribuído entre 3.707 classes distintas de assuntos, que funcionam como “escaninhos temáticos” em que são “encaixados” os itens (livros, teses, dissertações, periódicos, etc).

Este conjunto de quase 4.000 classes seria demasiado numeroso para análise. Assim, as variações do CDU foram consolidadas, tomando-se progressivamente menos dígitos no código, até que os ramos resultantes ficassem razoavelmente equilibrados em termos de número de ocorrência por ramo. Os ramos assim consolidados foram chamados de “Assuntos Significativos” (AS).

Em outras palavras, cada código CDU foi transformado em um código mais resumido, o AS. A Tabela 7, a seguir, exemplifica como foram consolidados os códigos CDU que resultaram nos Assuntos Significativos “33” e “330.1” dos itens com circulação no período do estudo.

Tabela 7**Exemplo de Consolidação de “Códigos CDU” em “Assuntos Significativos”**

Código CDU	Assunto Significativo	No. de Ocorrências
33	33	429
33(-77)	33	4
33(4)	33	5
33(420)	33	2
33(47)	33	4
33(510)	33	4
33(520)	33	31
33(6)	33	8
33(72)	33	11
33(73)	33	4
33(8=6)	33	47
33(81)	33	177
33(814.2)	33	10
33(815.1)	33	2
33(816.1)	33	2
33.001.1	33	4
330.101.52	330.1	9
330.101.541	330.1	373
330.101.542	330.1	241
330.115	330.1	76
330.123.7	330.1	5
330.123.7(81)	330.1	10
330.131.7	330.1	12
330.131.7(73)	330.1	4
330.14	330.1	2
330.14.011	330.1	6
330.146	330.1	14
330.146(81)	330.1	1
330.162	330.1	33
330.191.4	330.1	8
330.191.5	330.1	4

Após a recodificação, os 3.707 assuntos iniciais foram reduzidos a 42 assuntos significativos. As transações de circulação distribuem-se entre os assuntos significativos conforme a Tabela 8.

Tabela 8

Distribuição das Transações por Assuntos Significativos

Assunto Significativo	Descrição	Número de Transações
0	Generalidades. Ciência e Conhecimento. Organização	793
1	Filosofia. Psicologia.	648
2	Religião. Teologia.	24
3	Ciências Sociais. Direito. Administração, etc.	17
30	Metodologia.	180
31	Demografia. Sociologia.	532
32	Política.	601
33	Economia.	1,453
330.1	Teoria Econômica. Conceitos de Economia.	798
330.3	Dinâmica Econômica	422
331	Trabalho. Emprego. Economia do Trabalho.	393
334	Cooperativismo.	386
336	Finanças Públicas.	1,970
339	Comércio Internacional.	719
34	Direito. Jurisprudência.	1,552
35	Administração Pública. Governo. Assuntos Militares	309
36	Assistência, Previdência e Seguridade Social.	149
37	Educação. Ensino.	334
39	Antropologia.	110
5	Matemática. Ciências Naturais.	1,309
6	Ciências Aplicadas. Medicina. Tecnologia.	137
61	Ciências Médicas.	54
62	Engenharia. Tecnologia em Geral.	266
63	Agricultura.	61
65	Organização e Administração.	1,723
651	Escritório.	17
654	Telecomunicações e Telecontrole.	24
657	Contabilidade.	949
658	Administração de Empresas. Organização Comercial.	571
658.0	Administração.	335
658.1	Finanças.	808
658.3	Pessoal. Fator Humano. RH.	593
658.5	Administração da Produção.	863
658.6	Comércio.	98
658.7	Administração de Materiais.	185
658.8	Marketing, Vendas e Distribuição.	2,300
659	Publicidade e Propaganda. Relações Públicas.	219
68	Indústria.	213
7	Arte. Esportes.	102
8	Língua. Literatua.	114
9	Geografia. Biografia. História.	174
Total		22,508

Fonte: *Data Mart* de Circulação

2. IDENTIFICAÇÃO DE GRUPOS TEMÁTICOS

Uma vez criada a definição instrumental dos assuntos existentes na biblioteca (AS), passou-se a caracterizar o comportamento dos usuários com relação aos seus padrões de consultas temáticas. Decidiu-se empregar a técnica de Análise de Cesta (BERRY e LINOFF, 1997).

O nome da técnica evoca a imagem de um consumidor no supermercado, retirando produtos da prateleira e os colocando em sua cesta de compras. O conteúdo da cesta de um particular consumidor nos traz informação sobre seus hábitos de compra, característica familiares, etc. No outro extremo do processo de consumo, a análise do conteúdo da cesta de lixo de uma família também pode ser muito informativo e já foi freqüentemente utilizado como procedimento de controle ou investigação policial.

Se o conteúdo de uma única cesta é informativo, o conteúdo das cestas do conjunto dos clientes o é ainda mais. Pode, por exemplo (BERRY e LINOFF, 1997), fornecer uma boa noção de quais produtos são vendidos em conjunto, ajudando a identificar oportunidades de venda cruzada (*cross-selling*) .

Analogamente, na biblioteca, o conhecimento de quais assuntos significativos são consultados pelos mesmos usuários pode fornecer informação relevante sobre as características de sua pesquisa; e, as cestas do conjunto dos usuários pode fornecer informações sobre os itens mais relevantes em cada área temática.

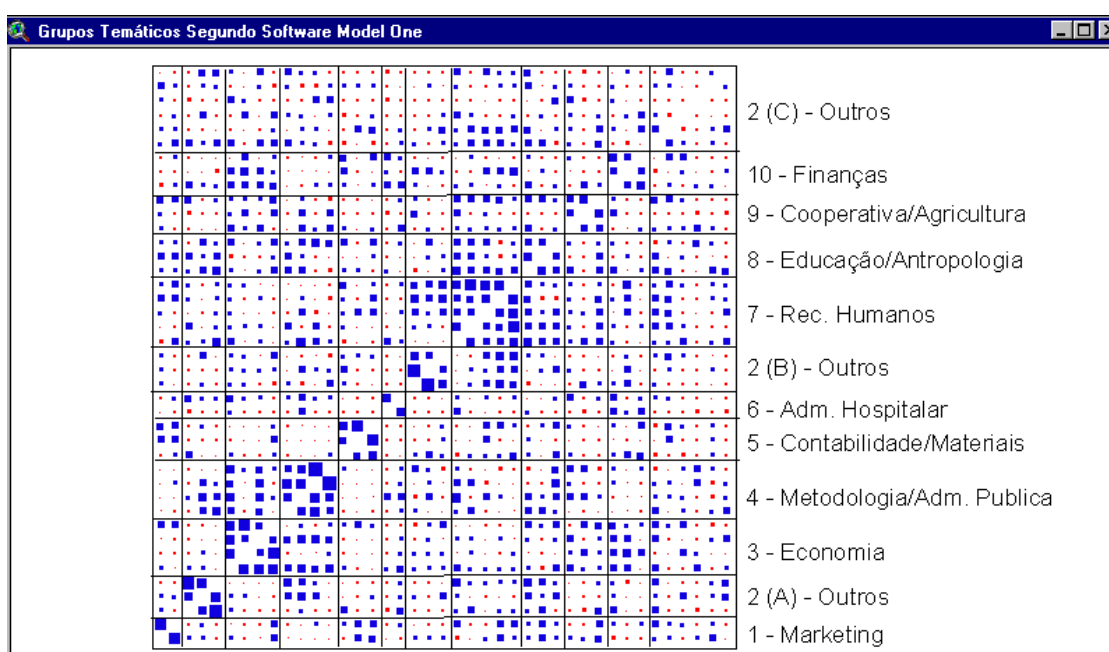
2.1. Análise de cestas com Model One

Com base nos pressupostos do item 2, os assuntos significativos dos itens emprestados no período focado foram estudados segundo a técnica de Análise de Cestas, utilizando-se o módulo de *cross-selling* do software Model One (UNICA,

1999). Os resultados obtidos estão representados na Figura 6, a seguir, correspondente à saída gráfica típica do Model One, reconstruída por meio do software ArcView (ESRI, 1999).

Figura 6

Análise de Cesta de Assuntos Significativos (Model One/ArcView)



Na matriz representada na figura, as linhas e colunas correspondem aos assuntos significativos, ordenados a partir do canto inferior esquerdo da matriz e segundo a mesma ordem em que aparecem na Tabela 9 a seguir. Linhas horizontais e verticais formando blocos delimitam grupos temáticos. Os quadros azuis são proporcionais ao grau de associação entre os assuntos significativos da linha e da coluna correspondentes (dois assuntos que frequentemente ocorrem juntos estão representados por quadros azuis grandes; dois assuntos que pouco ocorrem juntos

são representados por quadros azuis pequenos). A diagonal foi omitida, pois retrataria, por definição, a uma associação não informativa de 100%. Os quadros vermelhos são proporcionais a uma associação inversa (assuntos que tendem a inibir a ocorrência de outros estão representados por quadros vermelhos grandes se este efeito é forte; e pequenos, se é fraco). A legenda à direita da matriz indica o GT da faixa correspondente.

Os Grupos Temáticos foram definidos de forma que itens de um mesmo grupo tenham boa probabilidade de ocorrerem juntos; em consequência, os quadros azuis grandes ficaram concentrados em torno da diagonal. Este fenômeno é coerente com a hipótese, razoável, de que pesquisadores consultam muitos itens tematicamente correlatos.

Tabela 9**Grupos Temáticos e Correspondentes Assuntos Significativos, segundo Model One**

Grupo Temático	Desc. Grupo	Assunto	Desc. Assunto
1	Marketing	659	Publicidade e Propaganda. Relações Públicas.
		658.8	Marketing, Vendas e Distribuição.
2 (A)	Outros	8	Língua. Literatura.
		7	Arte. Esportes.
		2	Religião. Teologia.
3	Economia	339	Comércio Internacional.
		330.3	Dinâmica Econômica
		330.1	Teoria Econômica. Conceitos de Economia.
		33	Economia.
4	Metodologia/Sociologia/Política/Adm Pública	30	Metodologia.
		35	Administração Pública. Governo. Assuntos Militares
		32	Política.
		31	Demografia. Sociologia.
5	Contabilidade/Materiais	657	Contabilidade.
		658.6	Comércio.
		658.7	Administração de Materiais.
2 (B)	Outros	9	Geografia. Biografia. História.
		3	Ciências Sociais. Direito. Administração, etc.
6	Adm. Hospitalar	6	Ciências Aplicadas. Medicina. Tecnologia.
		61	Ciências Médicas.
		658.0	Administração.
7	RH	331	Trabalho. Emprego. Economia do Trabalho.
		1	Filosofia. Psicologia.
		658.3	Pessoal. Fator Humano. RH.
		65	Organização e Administração.
8	Educação/Antropologia	658	Administração de Empresas. Organização Comercial.
		0	Generalidades. Ciência e Conhecimento. Organização
		39	Antropologia.
9	Cooperativa/Agricultura	37	Educação. Ensino.
		R	Referência.
		68	Indústria.
10	Finanças	334	Cooperativismo.
		63	Agricultura.
		5	Matemática. Ciências Naturais.
2 (C)	Outros	336	Finanças Públicas.
		658.1	Finanças.
2 (C)	Outros	658.5	Administração da Produção.
		62	Engenharia. Tecnologia em Geral.
		36	Assistência, Previdência e Seguridade Social.
		654	Telecomunicações e Telecontrole.
		34	Direito. Jurisprudência.
651	Escritório.		

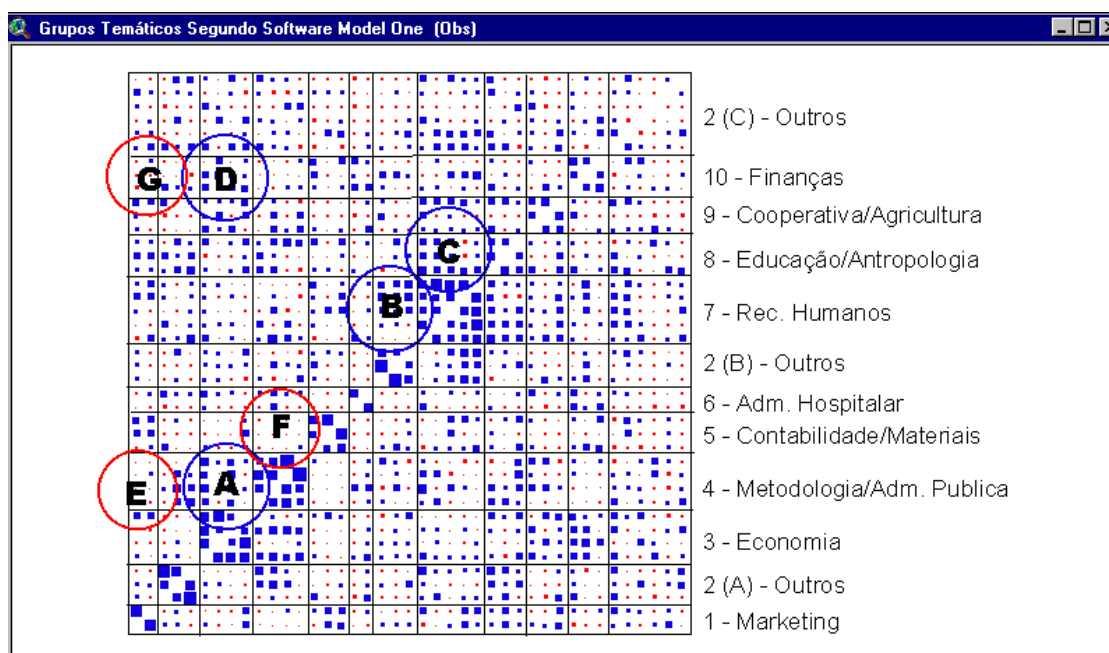
Os Grupos Temáticos foram nomeados de forma a refletir os assuntos significativos que contém. É fácil identificar nos Grupos Temáticos as principais áreas de atuação da Escola: Marketing, Finanças, Contabilidade, Administração de Materiais, Economia, Recursos Humanos, Administração Pública e Administração Hospitalar. O inverso, no entanto, não é verdadeiro: a partir das principais áreas de pesquisa da Escola, não seria fácil ou eficiente arbitrar como deveriam ser reunidos os assuntos significativos e, por conseqüência, as classes de assunto segundo o CDU.

Grupos Temáticos com um baixo número de ocorrências foram reunidos numa categoria “outros”. Estes GTs coincidem com tópicos pouco centrais na área de administração.

Note-se, na Figura 7 a seguir, mais pontos de coerência dos resultados obtidos com o que se conhece sobre as áreas de pesquisa da Escola: Administração Pública interage fortemente com Economia (Círculo A); Recursos Humanos têm muitos vínculos com temas marginais à Administração (Círculos B e C); Economia tem fortes laços com Finanças (Círculo D); por outro lado, Administração Pública quase nada tem a ver com Marketing e Contabilidade (Círculos E e F); e Finanças pouco interage com Marketing. O conhecimento das inter-relações entre os GTs é extremamente relevante para a montagem das listas de sugestões.

Figura 7

Algumas Considerações sobre os Resultados de Análise de Cesta (Model One/ArcView)



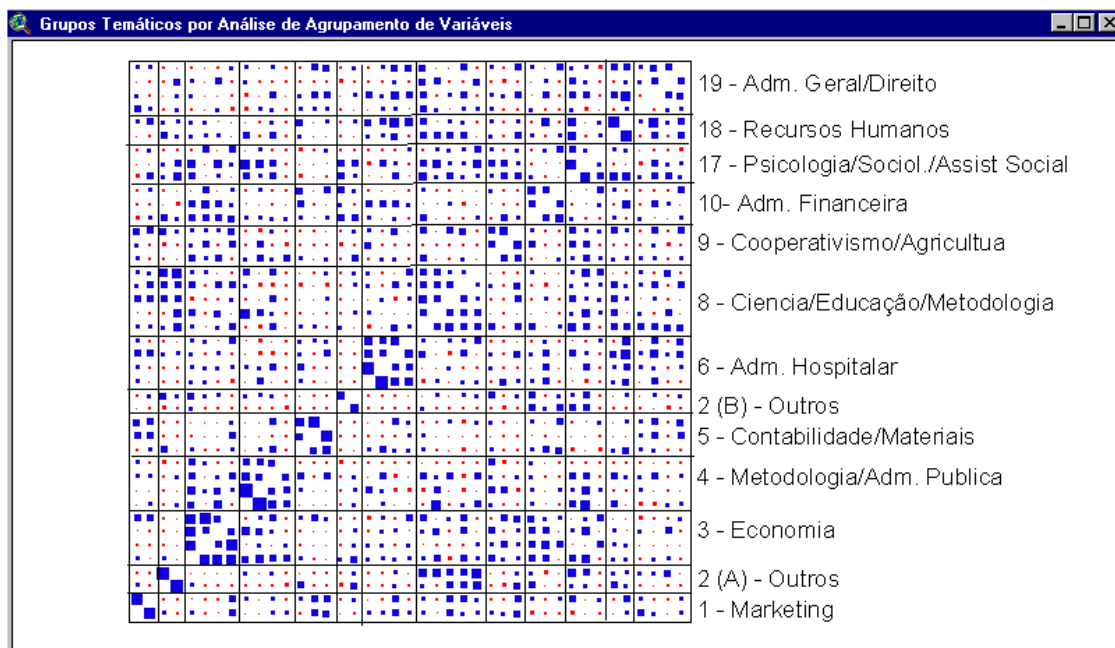
2.2. Validação da análise do Model One

Muito embora os resultados obtidos na análise de cesta com a aplicação do Model One tenham demonstrado coerência com a realidade da EAESP, decidimos validá-los estatisticamente, refazendo o estudo por meio de Análise de Agrupamentos (HAIR, ANDERSON e OUTROS, 1995; JACKSON, 1983).

Esta decisão tem duas justificativas:

- O software Model One é basicamente uma “caixa preta”; quase nada informa sobre os procedimentos que utiliza na derivação dos seus resultados o que é extremamente desconfortável do ponto de vista científico;
- A técnica de análise de agrupamento não exige software específico, podendo ser facilmente incorporada ao processo de carga dos dados no *data mart*, por meio de um procedimento armazenado.

Por conveniência de cálculo, na validação dos resultados do Model One foi utilizado o software Minitab. As conclusões da Análise de Agrupamento (obtidas com algoritmo de agrupamento de variáveis por correlação não absoluta e critério Ward de medida de distância) foram traduzidas para o padrão de apresentação do Model One e estão reproduzidas na Figura 8.

Figura 8**Análise de Cesta de Assuntos Significativos (Minitab/ArcView)**

Os resultados da validação são bastante consistentes com os obtidos pela aplicação do Model One, tendo os Grupos Temáticos principais coincido quase integralmente. Compare a Tabela 9 e a Tabela 10.

Tabela 10

Grupos Temáticos e Correspondentes Assuntos Significativos segundo Análise de Agrupamentos

Grupo Temático	Desc. Grupo	Assunto	Desc. Assunto
1	Marketing	659	Publicidade e Propaganda. Relacoes Publicas.
		658.8	Marketing, Vendas e Distribuicao.
2 (A)	Outros	8	Lingua. Literatua.
		2	Religiao. Teologia.
3	Economia	330.3	Dinâmica Economica
		339	Comércio Internacional.
		330.1	Teoria Economica. Conceitos de Economia.
		33	Economia.
4	Política/Administração Pública	32	Politica.
		35	Administração Publica. Governo. Assuntos Militares
		39	Antropologia.
		654	Telecomunicacoes e Telecontrole.
5	Contabilidade/Materiais	658.6	Comercio.
		657	Contabilidade.
		658.7	Administracao de Materiais.
2 (B)	Outros	9	Geografia. Biografia. Historia.
		3	Ciências Sociais. Direito. Administracao, etc.
6	Adm. Hospitalar	6	Ciências Aplicadas. Medicina. Tecnologia.
		658.0	Administracao.
		61	Ciências Medicas.
		658	Administracao de Empresas. Organizacao Comercial.
8	Ciência/Educação/Metodologia	0	Generalidades. Ciencia e Conhecimento. Organizacao
		30	Metodologia.
		37	Educacao. Ensino.
		R	Referencia.
		7	Arte. Esportes.
9	Cooperativa/Agricultura	62	Engenharia. Tecnologia em Geral.
		63	Agricultura.
		68	Industria.
10	Finanças	334	Cooperativismo.
		658.1	Financas.
		336	Finanças Publicas.
17	Psicologia/Sociologia/Assist.Social	5	Matematica. Ciências Naturais.
		36	Assistencia, Previdencia e Seguridade Social.
		1	Filosofia. Psicologia.
18	RH	31	Demografia. Sociologia.
		658.3	Pessoal. Fator Humano. RH.
19	Adm. Geral/Direito	331	Trabalho. Emprego. Economia do Trabalho.
		65	Organizacao e Administracao.
		651	Escritorio.
		658.5	Administracao da Producao.
		34	Direito. Jurisprudencia.

Na Tabela 10 foram listados em verde os Assuntos Significativos que na análise com o Minitab permaneceram no mesmo Grupo Temático em que haviam sido incluídos pelo Model One. Em vermelho, foram listados os Assuntos Significativos incluídos em GTs preexistentes. E em azul, os Assuntos Significativos remanejados para formar um Grupo Temático novo.

Diante da convergência de resultados, decidiu-se manter a classificação obtida com o software Model One, por facilidade de uso e por conveniência de nosso cronograma de trabalho.

2.3. Algumas considerações sobre os grupos temáticos

A Tabela 11 indica como as transações de circulação se distribuem entre os Grupos Temáticos, considerando-se todos os usuários da biblioteca e apenas os 410 usuários de Pós-Graduação e CEAHS.

No conjunto de todos os usuários, os GT com maior atividade são Finanças e RH. Entre os alunos de Pós+CEAHS são RH e Economia.

Note-se também como este último grupo tem um participação mais do que proporcional nos GTs de Educação/Antropologia e Metodologia/Adm. Pública. Uma investigação de dados mais detalhados do *data mart* revela que esta intensidade deve-se aos Assuntos Significativos **0 – Ciência e Conhecimento** (57% do uso geral) e **30 – Metodologia** (68% do uso geral). É razoável que os alunos de pós-graduação tenham um interesse mais acentuado por temas metodológicos do que alunos de graduação.

Tabela 11**Número de Transações por Grupo Temático**

GT	Descrição do GT	Transações		
		Todos os Usuários	Usuários de Pós + CEAS	%Linha
1	Marketing	2,519	969	38%
2	Outros	3,302	1,088	33%
3	Economia	3,392	1,286	38%
4	Metodologia/Adm. Pública	1,622	726	45%
5	Contabilidade/Materiais	1,232	334	27%
6	Adm. Hospitalar	526	210	40%
7	Recursos Humanos	3,928	1,732	44%
8	Educação/Antropologia	1,240	618	50%
9	Cooperativa/Agricultura	660	219	33%
10	Finanças	4,087	1,147	28%
	Total	22,508	8,329	37%

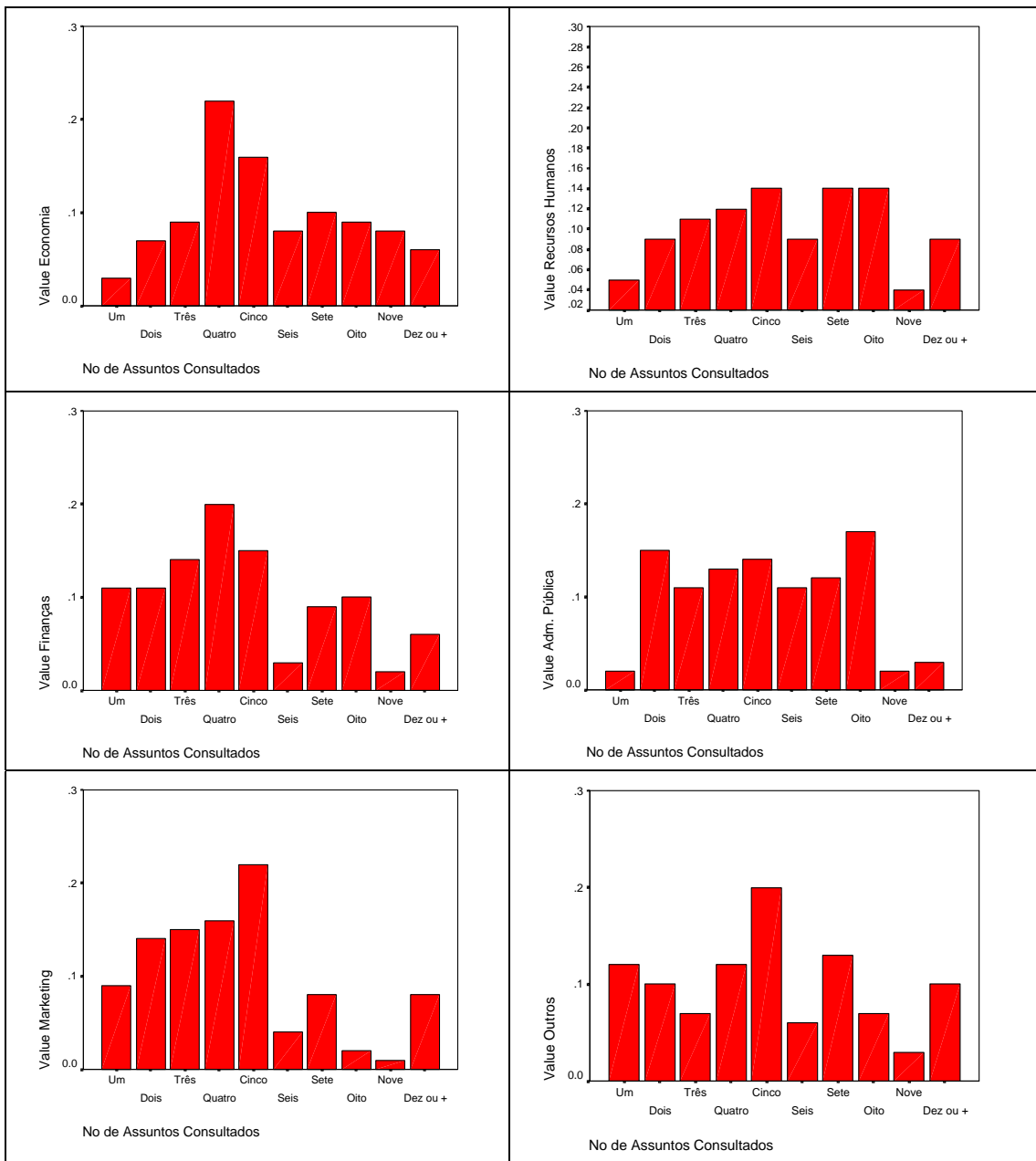
Fonte: *Data Mart* de Circulação

A análise do número de Assuntos Significativos diferentes pesquisados por usuários com atividade em cada grupo temático sugere a existência de Grupos Temáticos com diferentes perfis de usuários:

- **GTs muito especializados**, concentram usuários ativos em poucos assuntos distintos; isto acontece claramente nos Grupos Temáticos de Economia, Finanças e Marketing (coluna da esquerda na Figura 9);
- **GTs pouco especializados**, apresentam usuários distribuídos de forma razoavelmente uniforme entre as categoria de número de assuntos pesquisados (coluna da direita na Figura 9); isto acontece com os GT de Recursos Humanos, Administração Pública e Outros;
- **GTs Bimodais**, isto é, com uma concentração de usuários pesquisadores muito especializados e outra concentração de usuários pesquisadores pouco especializados; é o caso de Cooperativa/Agricultura e Educação/ Antropologia (vide Figura 10).

Figura 9

Perfil de Especialização dos Grupos Temáticos Focados e Genéricos

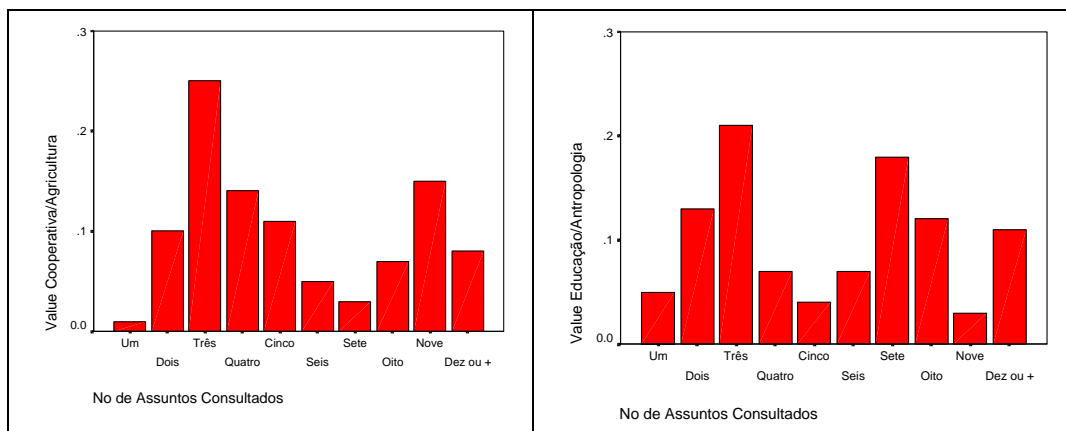


Nos dois primeiros casos, o perfil encontrado parece razoável: em nível de pós-graduação os temas de economia, finanças e marketing são, de fato, bastante especializados; já os temas de recursos humanos e administração pública parecem envolver uma maior gama de assuntos correlatos, ramificando-se por diversas área de saber.

Os grupos bimodais são mais difíceis de interpretar; talvez haja pesquisadores para os quais estes são temas centrais, enquanto para outros são temas acessórios; ou talvez esta configuração venha a se alterar conforme o *data mart* acumule dados de um período de tempo maior. Uma explicação mais conclusiva depende de estudos adicionais que fogem ao escopo deste trabalho.

Figura 10

Perfil de Especialização de Grupos Temáticos com Distribuição Bi-Modal



3. CRIAÇÃO DOS SUBGRUPOS ESPECIALIZADOS DE PESQUISADORES PARA CADA GRUPO TEMÁTICO

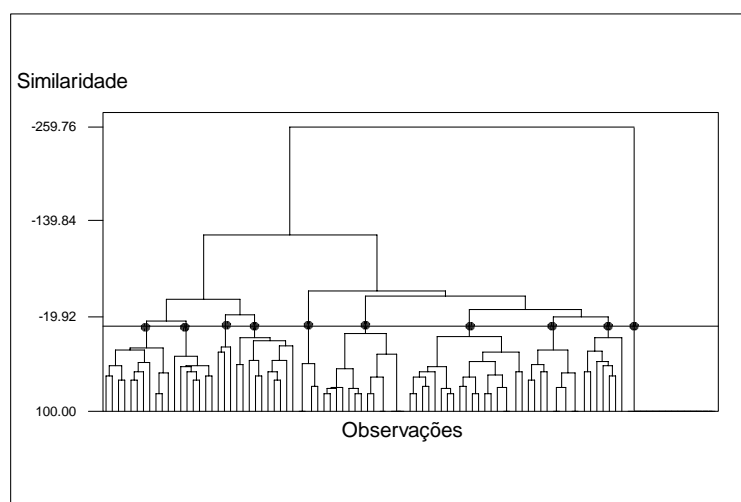
Uma vez definidos os Grupos Temáticos e os Assuntos Significativos (que são atributos dos itens do acervo) pode-se utilizá-los para caracterizar Subgrupos Especializados de Usuários. Para fazê-lo, propõem-se as etapas descritas a seguir.

3.1. Separação dos usuários com transações em um GT de interesse

Uma consulta ao *data mart* permite separar os usuários que fizeram transações envolvendo um GT, digamos, para exemplificação, o Grupo Temático de Marketing. Nesta situação temos 99 dos 410 usuários estudados. Para efeito de segmentação, estes usuários passam a ser considerados em separado. O processo é repetido para os demais GTs, um de cada vez.

3.2. Agrupamento dos usuários conforme similaridade no perfil de assuntos significativos

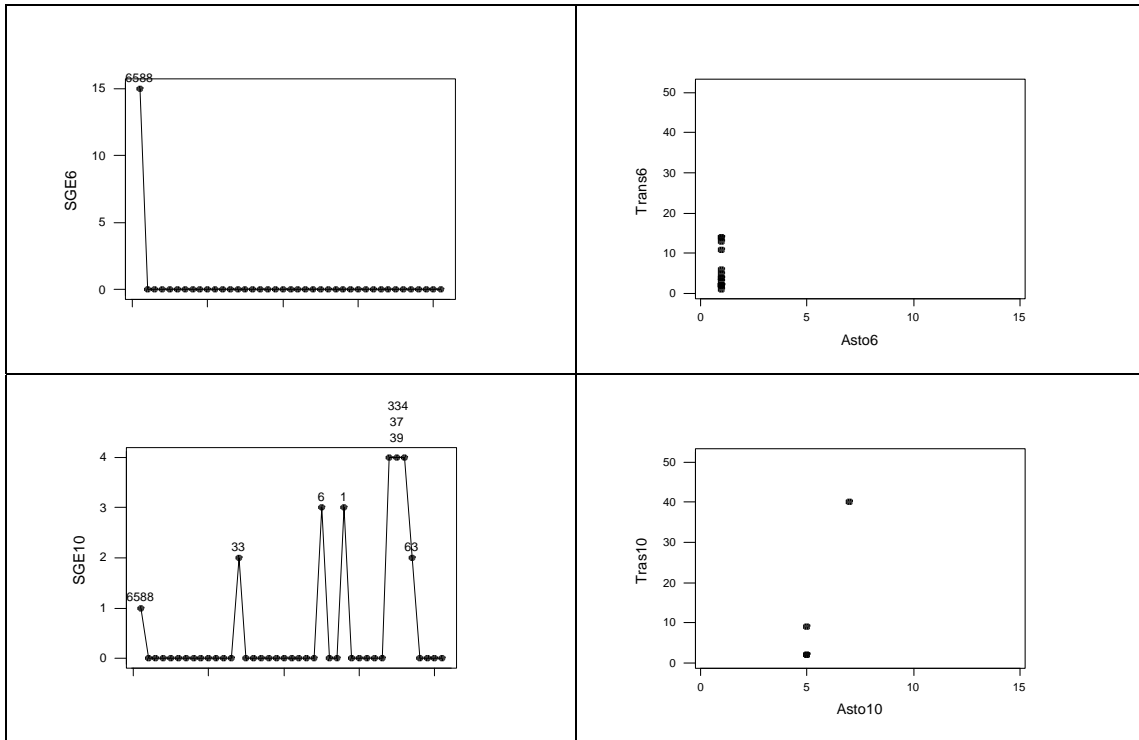
Os usuários do GT em foco, em nosso exemplo os usuários com transações no GT de Marketing, são, a seguir, separados em grupos homogêneos com relação ao seu perfil de consulta aos 42 Assuntos Significativos definidos em IV.1. Para isso foram utilizados algoritmos de Análise de Agrupamento; depois de várias tentativas exploratórias, adotamos um algoritmo hierárquico aglomerativo com distância euclidiana e critério Ward. O dendrograma obtido está representado a seguir, na Figura 11. Com base nos resultados da análise deste dendrograma, cada usuário foi alocado a um Subgrupo Especializado, numerado de 1 a 10.

Figura 11**Dendrograma dos Usuários com Transações no GT de Marketing**

Nos gráficos da Figura 12 até a Figura 15, os subgrupos encontrados na análise de agrupamento de usuários foram caracterizados de duas maneiras:

- **a coluna da esquerda** destes gráficos representa o perfil de Assuntos Significativos consultados pelos membros do grupo; no SGE 6 (Figura 12), por exemplo, todos os usuários consultaram exclusivamente o assunto 658.8 – Marketing, Vendas e Distribuição; já no SGE 9 (Figura 14), além do assunto 658.8, os usuários tinham um foco secundário no assunto 33 – Economia, que se destaca contra um pano de fundo bastante diversificado;
- **a coluna da direita** das figuras mostra o grau de especialização do grupo em termos do número de assuntos consultados (eixo horizontal) *versus* o nível de atividade do grupo em termos do número de transações realizadas por seus membros; os SGEs retratados na Figura 13 são mais focados do que os representados na Figura 14.

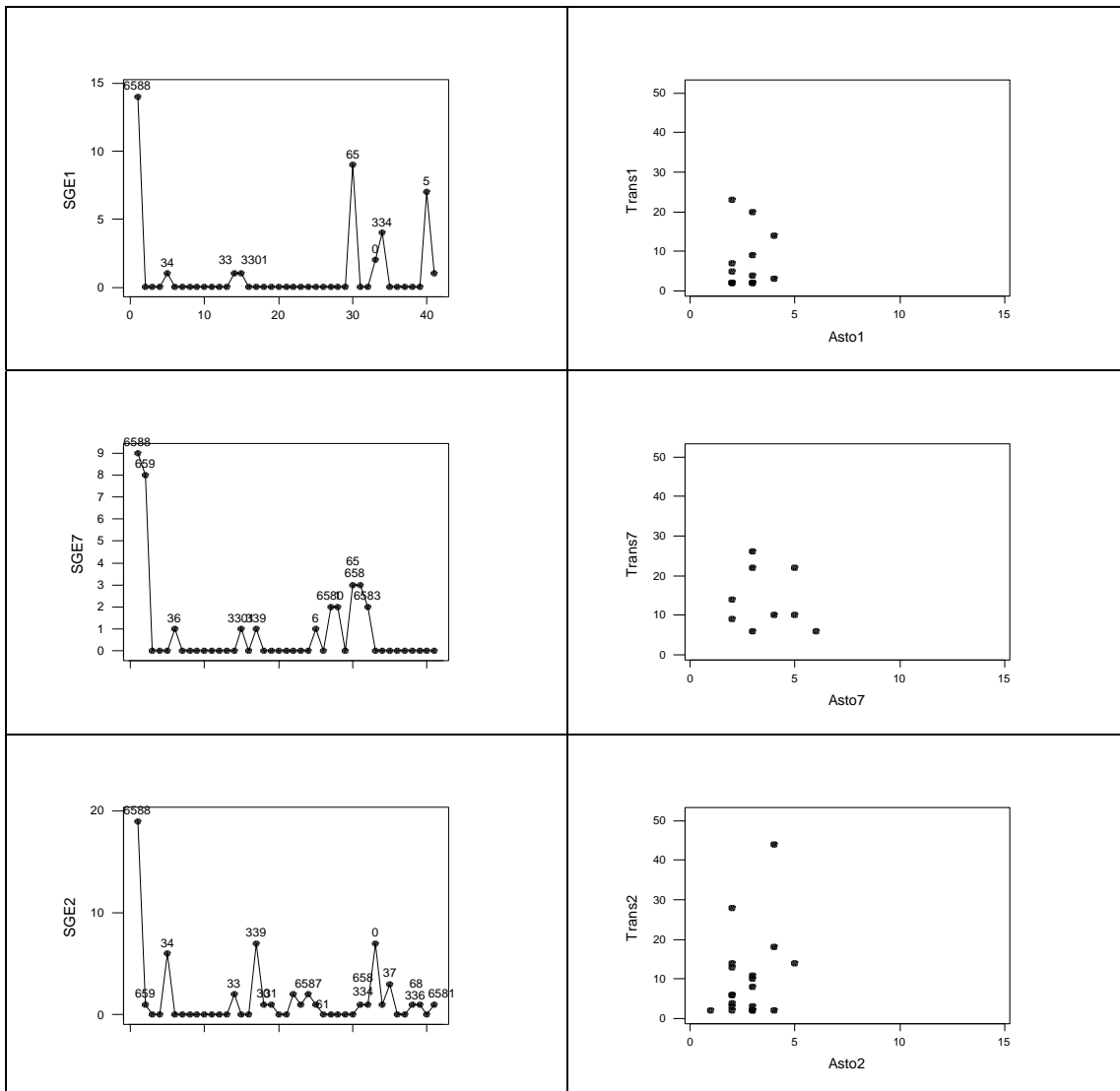
Figura 12

Assunto Central *versus* Assunto Marginal

Na Figura 12, procuramos contrastar o SGE 6, em que obviamente Marketing é assunto focal, com o SGE 10, em que Marketing é um assunto marginal – o foco está em Cooperativismo e Antropologia. O tipo de recomendação a ser feita, em cada caso, deve ser diferente, pois o assunto do GT tem significados distintos para estes SGEs.

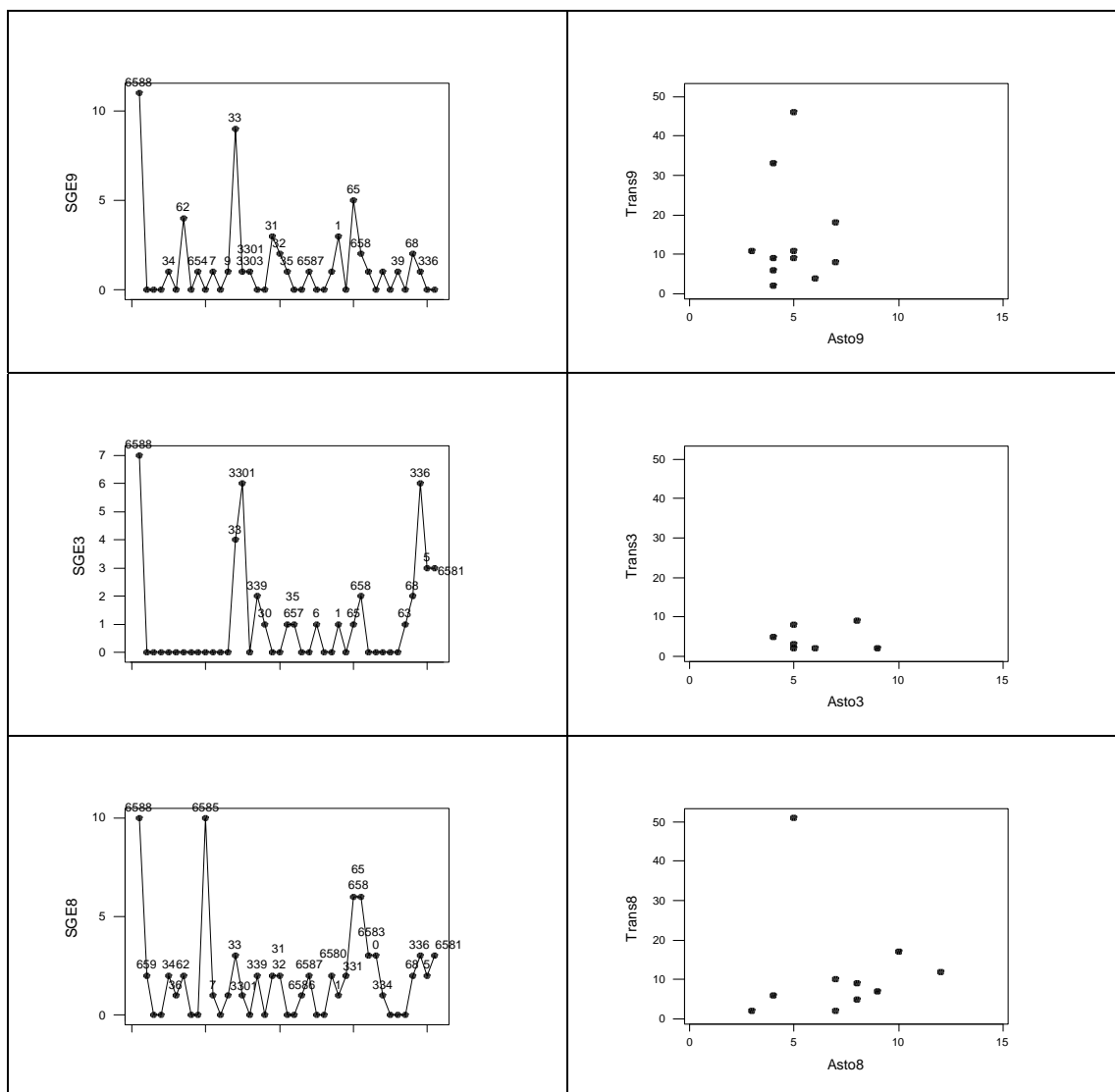
Figura 13

Grupos Muito Especializados



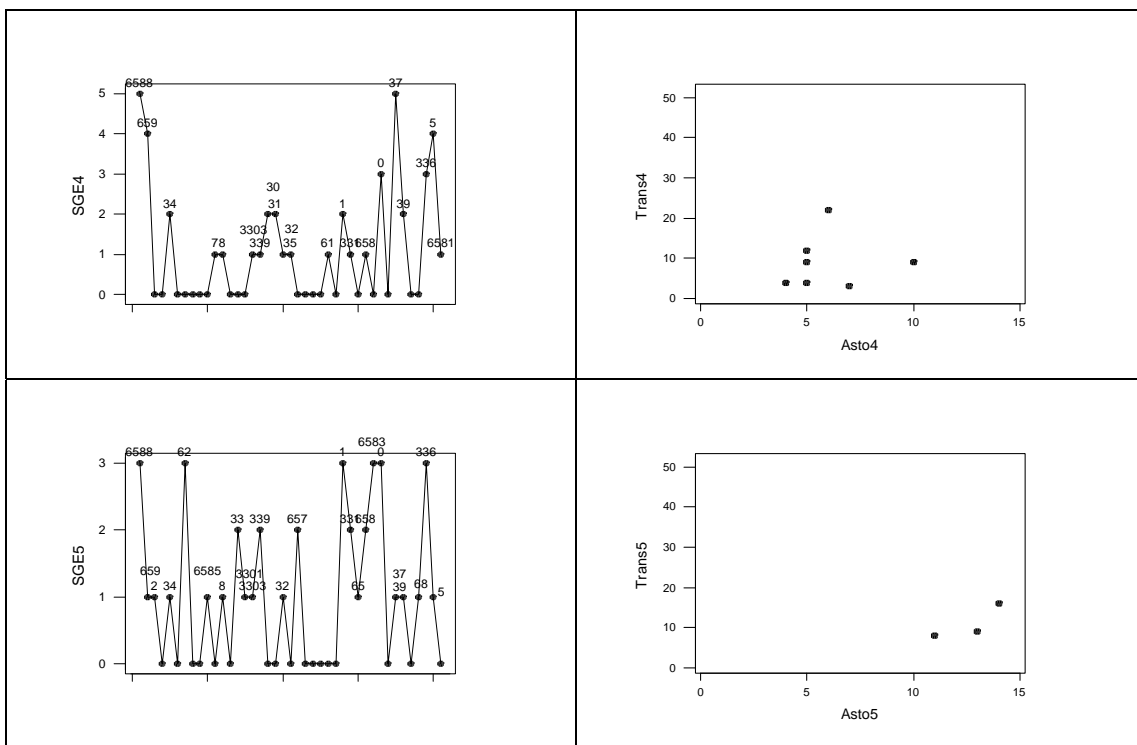
A Figura 13 reuniu os grupos mais especializados, com exceção do SGE 6, também especializado, mas que apareceu na Figura 12.

Figura 14
Grupos Pouco Especializados



A Figura 14 reuniu os SGEs pouco especializados. Na Figura 15 foram representados os SGEs muito diversificados.

Figura 15
Grupos Muito Diversificados



4. CRIAÇÃO DE LISTAS-BASE PARA O SGE

Uma vez identificados os Subgrupos Especializados de um Grupo Temático, a próxima etapa na elaboração das recomendações personalizadas é a consolidação das transações dos membros do SGE de forma a produzir uma Lista-Base do SGE (LBSGE).

Continuando com nosso exemplo, em que o GT de Marketing foi segmentado em 10 SGEs, analisemos agora o SGE 10. A consolidação das transações dos membros deste Subgrupo Especializado resultou na Tabela 12, a seguir.

Tabela 12

Lista-Base do SGE 10 do GT 1- Marketing

CDU	Item Título	Ranking	Patron Key			
			410	3152	3194	5953
658.62	Gerenciamento por categorias: melhores praticas	4	2	2	2	13
658.818	ECR Brasil : visao geral e potencial de redução de custos	3	2	2		7
658.788	Padronização	3	2	2		9
658.62	Reposição continua de mercadorias	3	2	2		10
657.47	Custeio baseado em atividades	3	2	2		8
657.422.2	Processos financeiros	3	2	2		9
65.012.45	EDI aplicado a cadeia de abastecimento	3	2	2		6
659.126.1	Retail power plays: from trading to brand leadership : s...	2			2	5
658.8	Trade marketing strategies: the partnership between manu...	2			2	9
659.126.1	Marcas de supermercado	1				3
658.86/.87	Retail distribution management: a strategic guide to dev...	1				9
658.86/.87	Marketing channels: a management view	1				5
658.8	Marketing de relacionamento: estratégias bem sucedidas	1			2	
658.8	Databse marketing estrategico	1			2	
658.8	Tecnologia de informação e comunicação	1				2
658.8	Principios de marketing	1				1
658.7	Retail buying	1				2
658.7	Políticas de suprimento, tecnologia de producao e tecnologia	1		1		
339.37	Contemporary retailing	1				2
339.37	Retail saturation :examining the evidence	1				5
339.37	Retail management	1				2
339.37	Retailing management	1				2
339.37	Modern retailing: theory and practice	1				2
339.37	Retailing: new perspectives	1				2

Na Tabela 12, “Patron Key” é a identificação do usuário no *Data Mart* de Circulação; “Ranking” corresponde ao posto do item, tendo sido os itens ordenados decrescentemente segundo o número de usuários que o consultaram, sendo o máximo possível 4, e o mínimo, 1. Os valores registrados nas colunas correspondentes aos usuários representam o número de transações realizadas por eles, envolvendo um mesmo item.

A interpretação da tabela nos revela que:

- O SGE 10 é formado por 4 usuários da biblioteca;

- No total, estes 4 usuários levaram emprestados 24 itens, dos quais:
 - um item foi consultado por todos os usuários;
 - seis foram consultados por 3 entre os 4 usuários;
 - dois itens foram consultados por dois usuários;
 - e 15 itens foram consultados apenas por um dos usuários;
- Nem todos os itens são estritamente sobre marketing, mas são correlacionados ao tema;
- Os itens parecem formar um conjunto tematicamente coerente;
- O usuário “5953” costuma renovar repetidamente o empréstimo dos itens que consulta, enquanto os demais os retiram apenas uma vez (realizando, portanto, duas transações, uma de empréstimo, outra de devolução).

5. CRIAÇÃO DE LISTAS TEMÁTICAS

Uma lista com indicações menos especializadas é elaborada simplesmente selecionando-se os, digamos, 20 itens que mais circularam dentro de um GT. No caso de Marketing são:

Tabela 13**Lista de Recomendações Temáticas: SGE 1 – Marketing**

CDU	Título	Transações
658.8.012.l	Marketing research	28
658.89	Consumer behavior	26
658.8	Services marketing	25
658.8.012.1	Pesquisa de marketing : metodologia, planejamento, execu,	24
62::007	Tecnologia de informação: o uso de TI pelas empresas...	22
303.4	Métodos de pesquisa nas relações sociais Selltiz, Wrigh...	22
658.62	Gerenciamento por categorias: melhores práticas...	19
62::007	Gerência estratégica da tecnologia da informação	17
658.89	The service profit chain: how leading companies link pro...	16
519.233.5	Análise de variância e de regressão : uma introdução	16
338.46	Serviços revolucionários: mudando as regras do jogo com...	16
336.763	An introduction to the mathematics of financial derivatives	15
658.8	Administração de marketing	15
658.8	Marketing de serviços: conceitos e estratégias	14
658.8	Industrial marketing management: text and cases	14
292	Mitologia	14
658.62	Reposição contínua de mercadorias	14
657.47	Custeio baseado em atividades	14
658.8	Fundamentos de marketing	14
338.46	Service management: operations, strategy, and informatio	14

6. CRIAÇÃO DE UMA LISTA GERAL

Finalmente, uma lista de itens de interesse geral pode ser elaborada, selecionando-se os itens do acervo que mais circularam no conjunto dos usuários. Os 20 primeiros são:

Tabela 14**Lista de Recomendações Gerais**

CDU	Título	Transações
330.101.541	Macroeconomia	159
658.15	Princípios de administração financeira	146
657.622	Análise financeira de balanços: abordagem básica e gere...	133
336.76(81)	Mercado financeiro : produtos e serviços	118
658.8	Administração de marketing: análise, planejamento, im...	115
658.8	Princípios de marketing	97
342(81)	Curso de direito constitucional	95
519.2	Statistics for business & economics	92
519.2	Estatística básica	87
658.15	Princípios de administração financeira	86
658.3	Recursos humanos	84
519.2	Estatística	81
65.01	Imagens da organização	76
658.8.012.1	Pesquisa de marketing : metodologia, planejamento, execu...	76
657	Contabilidade empresarial	75
657.012.4	Contabilidade para administradores	74
657	Contabilidade introdutória	73
658.5	A meta: um processo de aprimoramento contínuo	69
34::336.2(81)	Curso de direito tributário	69
330.101.542	Microeconomia	67

Observe-se como estes itens tendem a refletir os livros didáticos mais utilizados pelos alunos de graduação. São recomendações com pouco valor agregado para os usuários.

7. CRIAÇÃO DA LISTA PERSONALIZADA

Para a criação das recomendações personalizadas, procede-se a uma consolidação dos listas especializada, temática e geral, com maior ênfase na lista especializada.

Digamos que se pretenda exibir 10 recomendações: sugerimos utilizar 6, 3 e 1 itens respectivamente de cada fonte.

7.1. Seleção de itens da lista especializada

Selecionam-se da lista especializada os itens de maior ranking e que ainda não foram tomados emprestados pelo usuário. Assim, por exemplo o usuário com Patron Key 410 (vide Tabela 12) receberia as sugestões seguintes:

Tabela 15

Sugestões Especializadas Personalizadas para o Usuário 410

Item		
CDU	Título	Ranking
659.126.1	Retail power plays: from trading to brand leadership : s...	2
658.8	Trade marketing strategies: the partnership between manu...	2
659.126.1	Marcas de supermercado	1
658.86/.87	Retail distribution management: a strategic guide to dev...	1
658.86/.87	Marketing channels: a management view	1
658.8	Marketing de relacionamento: estratégias bem sucedidas	1

8. EXEMPLOS DOS RESULTADOS OBTIDOS

O resultado final de uma lista de recomendações para o usuário 410 poderia tomar a forma das figuras a seguir.

Figura 16

Sugestões Personalizadas

Olá, Francisco. Nós achamos que estes itens podem ser do seu interesse:

- **Itens específicos para o seu perfil na área de [Marketing](#)**

658.8 M478ma Regis McKenna	Marketing de Relacionamento: Estratégias Bem Sucedidas para a Era do Cliente. Rio de Janeiro: Campus, c1997
658.8 R188t Geoffrey Randall.	Trade Marketing Strategies: The Partnership Between Manufacturers, Brands and Retailers. Oxford: Butterworth Heinemann, 1994
658.86/.87 R437 John Fernie (org.)	Retail Distribution Management: A Strategic Guide to Developments and Trends. London: Kogan, 1990.
658.8 R815m Bert Rosenbloom	Marketing Channels: A Management View. Fort Worth: Dryden, c1999
659.126.1 P436ma Inês Pereira	Marcas de Supermercado. EAESP/FGV Tese 1998.
659.126.1 W676r Andrew Wileman	Retail Power Plays: From Trading to Brand Leadership. Basingstoke: Macmillan, 1997

[Mais sugestões especializadas em Marketing](#) [Refinar sugestões](#)

- **Itens gerais na área de [Marketing](#)**

658.89 E57c James F. Engel	Consumer Behavior Forth Worth : Dryden, c1995
658.89 W418m Ronald M. Weiers	Marketing Research Englewood Cliffs : Prentice-Hall, c1984
658.8 L911s Christopher Lovelock	Services Marketing Englewood Cliffs : Prentice-Hall, c1991

[Mais sugestões gerais em Marketing](#) [Refinar sugestões](#)

Figura 16

Sugestões Personalizadas (continuação)

- **Item popular entre os usuários em geral**

<u>330.101.541 S611m</u> <u>Mário H. Simonsen</u>	Macroeconomia Rio de Janeiro : Ao Livro Tecnico, 1989
--	--

[Mais sugestões de itens populares entre os usuários](#) [Refinar sugestões](#)

Observe que nem mesmo as recomendações gerais serão as mesmas para diferentes usuários, uma vez que recomendações já retiradas em empréstimo por ele são eliminadas de sua lista. Elas também o são após um determinado número de exibições na página.

9. TRANSAÇÕES VIRTUAIS

Quando o usuário opta por “Refinar suas sugestões”, conforme opção de *link* sugerida na Figura 16, o sistema deve dar-lhe a oportunidade de classificar os itens sugeridos numa escala do tipo:

- Conheço o item e ele é irrelevante para mim;
- Não conheço o item mas ele me parece irrelevante;
- Conheço o item e ele é relevante para mim;
- Conheço o item e ele é muito relevante para mim.

Esta informação fornecida pelo usuário tem dois efeitos: gera um “registro de transação virtual”, em tudo análogo ao de uma transação real, que será incorporado

ao *data mart* e usado no preparo de novas listas para o usuário; e provoca a eliminação do item da lista de recomendações, para que seja substituído por outro.

10. PRÓXIMOS PASSOS

A metodologia de identificação de perfil de usuários e geração de listas personalizadas de recomendações desenvolvido neste projeto de pesquisa está, a menos de pequenos ajustes, pronta para ser colocado em prática no contexto da Biblioteca Karl Boedecker; o *data mart* necessário está desenhado e suas rotinas de carregamento e atualização estão escritas em SQL; resta verificar a viabilidade econômica e tecnológica de se criar o *front end* necessário para realizar transações eletrônicas pela Internet.

O desdobramento deste projeto em um piloto operacional dependerá, entre outras coisas, da identificação de patrocinadores para a implantação da tecnologia de *web* necessária. Considerando-se a aplicabilidade de sistemas de recomendações como o aqui sugerido em praticamente todas as áreas de comércio eletrônico, imaginamos que isto será possível. Temos interesse especial por testar a tecnologia que, a partir das informações disponibilizadas no *Data Mart* de Circulação, **personaliza** as páginas exibidas pelo sistema ao usuário.

V. OPORTUNIDADES DE PESQUISA

Ao longo deste projeto, foram identificados vários temas que podem ser desenvolvidos para aperfeiçoar a inteligência do Sistema de Recomendações. Parece-nos particularmente desejável uma investigação sobre a teoria e prática da Análise de Redes para a descoberta de oportunidades de cooperação indireta (SCHWARTZ e WOOD, 1993; WASSERMAN e FAUST, 1997) e sobre a

representação gráfica destas redes (FREEMAN, 1999). Este tipo de análise requer a disponibilidade de software específico, que precisam ser adquiridos e estudados.

Uma outra linha de estudos interessante seria validar a relevância das recomendações geradas pelo sistema; para isso poderia ser realizada uma pesquisa junto aos usuários para os quais foram geradas as listas de sugestões. Para isso seria antes necessário o desenvolvimento de uma métrica adequada, possivelmente envolvendo os postos dos itens recomendados na listagem geral e na listagem especializada.

VI. CONCLUSÕES FINAIS

Neste projeto procuramos demonstrar como, a partir de um *data mart* de transações de circulação do acervo, uma biblioteca pode, implementando um sistema de cooperação indireta, gerar recomendações de leitura que criam valor aos seus usuários pesquisadores.

O sistema de geração das listas personalizadas de recomendações pode funcionar baseado em procedimentos de consolidação e conexão de registros em um *data mart*, associados às técnicas de Análise de Cestas e Análise de Agrupamentos.

Foram definidos os conceitos de Assuntos Significativos (baseados em atributos de itens) e de Grupos Temáticos (baseado no comportamento dos usuários); em seguida foram definidos Subgrupos Especializados de usuários (com base em seu perfil de leitura). Os usuários foram alocados nos segmentos identificados. Finalmente foram geradas as listas de sugestões, contendo recomendações específicas, temáticas e gerais. Sugeriu-se a utilização de transações virtuais para aperfeiçoar as recomendações.

O Sistema de Recomendações proposto pode ser implementado a baixo custo; a implementação por Internet depende apenas da disponibilização de ferramentas de

web para exibição de páginas personalizadas; é aplicável a bibliotecas e livrarias virtuais; pode ser adaptado a praticamente qualquer tipo de aplicação de *e-business*.

VII. BIBLIOGRAFIA

BERRY, Michael e LINOFF, Gordon. *Data Mining Techniques For Marketing, Sales and Customer Support*. New York: John Wiley and Sons, 1997, 454 pp.

BERSON, Alex e SMITH, Stephen. *Data Warehousing, Data Mining & OLAP*. New York: MacGraw-Hill, 1977, 612 pp.

BIBLIOTECA KARL A. BOEDECKER. *Relatório Estatístico Anual de 1998, documento interno*.

CABENA, Peter; HADJINIAN, Pablo; e outros. *Discovering Data Mining: From Concept to Implementation*. Upper Saddle River: Prentice Hall, 1998, 195 pp.

CARSON, Kerry D; CARSON, Paula P. e PHILIPS Joyce S. *The ABCs of Collaborative Change: The Manager's Guide to Library Renewal*. Chicago: ALA Editions, 1997, 272 pp.

DONNARD, Heloisa e BOREGAS, Edmilson. *A Experiência da Biblioteca no Processo de Informatização*. Documento apresentado no 1º Encontro Nacional de Usuários do VTLS, realizado na UERJ, Universidade Estadual do Rio de Janeiro, RJ, 16 de Setembro de 1.998.

DYER, C; BROWN, R; e GOLDSTEIN, D. *School Libraries: Theory and Practice*. London: Clive Bingley, 181 pp.

ECO, Humberto. *Como se Faz uma Tese em Ciências Humanas*. Lisboa: Presença, 1982.

ESRI. *Geography Matters*. <http://www.esri.com>, link válido em 05/01/00 às 15h17.

FREEMAN, Linton. *Visualizing Social Networks*. <http://carnap.ss.uci.edu/vis.html>, link válido em 27/12/99 às 9h45.

GOLDBERG, Henry e SENATOR, Ted. “*Restructuring Databases for Knowledge Discovery by Consolidation and Link Formation*”. Proceedings of the First International Conference on Knowledge Discovery and Data Mining. Menlo Park: AAAI Press, 1995.

HAIR, Joseph; ANDERSON, Rolph; e outros. *Multivariate Data Analysis*, 4th ed., Englewood Cliffs: Prentice Hall, 1995, 745pp.

HANN, Leslie. “*High-Tech Sleuths*”. Best’s Review, Nov. 1998, pp. 83-85.

IBICT. *Classificação Decimal Universal*. Brasília: MCT/CNPQ/IBICT, 1987, 505 pp.

JACKSON, Barbara. *Multivariate Data Analysis: An Introduction*. Homewood: Irwin, 1983, 242 pp.

JENSEN, David. “*Prospective Assessment of AI Technologies for Fraud Detection*”. Working Papers of the AAAI-99 Workshop on Artificial Intelligence Approaches to Fraud Detection and Risk Management.

KAUTZ, Henry; SELMAN, Bart; e SHAH, Mehul. “*Combining Social Networks and Collaborative Filtering*”, Communications of the ACM, March 1997, vol. 40, n. 3, pp. 63-65.

- KAUTZ, Henry; SELMAN, Bart; e SHAH, Mehul. “*The Hidden Web*”, AI Magazine, Summer 97, pp. 27-36. Disponível em <http://www.research.att.com/~kautz/referralweb/doc/aimag.pdf>, link válido em 27/12/99 às 11h10).
- KIMBALL, Ralph. *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. New York: John Wiley & Sons, 1996.
- KOTLER, Philip. “*Personalização em Massa*”, HSM, novembro/dezembro 1997, pp. 136-144.
- MILLER, Kathy. *Librarians Getting a Handle on Knowledge*. Information World Review, June 1998, pp. 25-26.
- PAYTON, David. “*Dicovering Collaborators by Analyzing Trails Through an Information Space*”, Artificial Intelligence and Link Analysis Papers from the 1998 Fall Symposium, October 23-25, Orlando, Florida.
- PEPPERS, Don e ROGERS, Martha. “*Is your company ready for one-to-one marketing?*”, HBR, January-February 1999, pp. 151- 163.
- RAMO, Joshua C. “*The Fast-Moving Internet Economy Has A Jungle of Competitors... And Here Is the King: Jeffery Preston Bezos, 1999 Person of the Year*”, TIME Magazine, 27/12/1999, pp. 42-59.
- SANTOS, Érico R. *Implantação de Tecnologia de Data Warehouse em Bibliotecas com Uso de Tecnologia Adequada*. São Paulo:EAESP/FGV, Relatório Parcial de Projeto de Pesquisa do PIBIC – Programa de Bolsas de Iniciação Científica, 1999, 53 pp.
- SILVA, Odilon. *Manual de CDU*. Brasília: Briquet de Lemos, 1994, 89 pp.

SCHWARTZ, Michael e WOOD, David. “*Discovering Shared Interests Using Graph Analysis*”. Communications of the ACM, August 1993, vol. 36, n. 8, pp. 78-89.

SWANSON, Don e SMALHEISER, Neil. *Link Analysis of Medline Titles as an Aid to Scientific Discovery*. <http://kiwi.uchicago.edu/libtrends.html>, link válido em 04.01.00 às 11h41.

UNICA. *Intelligent Solutions for the Business of Marketing*. <http://www.unica-usa.com>, link válido em 05/01/00 às 15h20.

VTLS INC. *Library Automation*. <http://www.vtls.com>, link válido em 26/12/99 às 16h53.

WASSERMAN, Stanley e FAUST, Katherine. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994, 825 pp.

WURMAN, Richard. *Ansiedade de Informação*. São Paulo: Cultura, 1991. 380 pp.