

Lekce 3 - Data v GIS

1. Cíle lekce.....	1
2. Geografická data.....	1
3. Rastrový a vektorový model prostorových dat.....	2
4. Data v rastrovém modelu.....	4
5. Data ve vektorovém modelu.....	4
5.1 Špagetový (spaghetti) model prostorových dat.....	4
5.2 Topologický model prostorových dat.....	5
6. Základní entity prostorových dat.....	5
6.1 Geometrické entity	5
6.2 Topologické entity.....	6
7. Vztahy mezi prostorovými daty.....	8
8. Zkušební otázky.....	8

1. Cíle lekce

- vysvětlit pojem geograficky vztažená data, prostorová data
- popsat vektorový a rastrový model prostorových dat
- vymezit základní entity prostorových dat a vztahy mezi nimi

2. Geografická data

Geografická proměnlivost reálného světa je nekonečně komplexní, k jejímu přesnému zachycení bychom potřebovali nekonečně velkou databázi. Data tedy musí být redukována na konečnou a spravovatelnou velikost. Pravidla, která jsou užitá pro konverzi reálného světa do diskretních objektů vytvářejí datový model.

Datové modelování v GIS se neliší od modelování běžných IS. Datový model každého informačního systému obsahuje objekty (entity) a relace mezi objekty (entitami)¹. Mezi relacemi je významná relace hierarchie (vyjadřuje nadtyp - podtyp objektu), tato relace je acyklická², nemusí to být strom (les). V datovém modelu GIS jsou vymezeny speciální - prostorové - objekty a speciální - prostorové - relace. Prostorové objekty a relace jsou pokud možno zobecňovány a standardizovány: architekt systému vybírá z předem stanovených tříd prostorových objektů a z předem stanovených tříd prostorových relací.

Prostorové objekty se dále dělí dále na dva základní podtypy

- geometrický objekt, který popisuje tvar objektu a lokalizaci objektu v prostoru - například souřadnicemi
- topologický objekt, který vchází do prostorové relace s jinými (topologickými) objekty

Neprostorové objekty se v GIS nazývají předmětné objekty. Všechny objekty a relace jsou v každém datovém modelu definovány svým výčtem. Například:

- I. prostorový objekt
 - A. prostorový topologický objekt
 1. třída bod
 2. třída graf
 - a) třída uzel
 - b) třída hrana
 3. třída hranice areálu
 4. třída areál
 - B. prostorový geometrický objekt
 1. třída vektorových dat
 - a) třída bod

¹ Omezíme se na binární relace na množině všech objektů obsažených v datovém modelu. Relace mezi objekty (entitami) mohou mít různé názvy v závislosti na použité metodice. Například v návrzích objektově orientovaných systémů se často nazývají asociace.

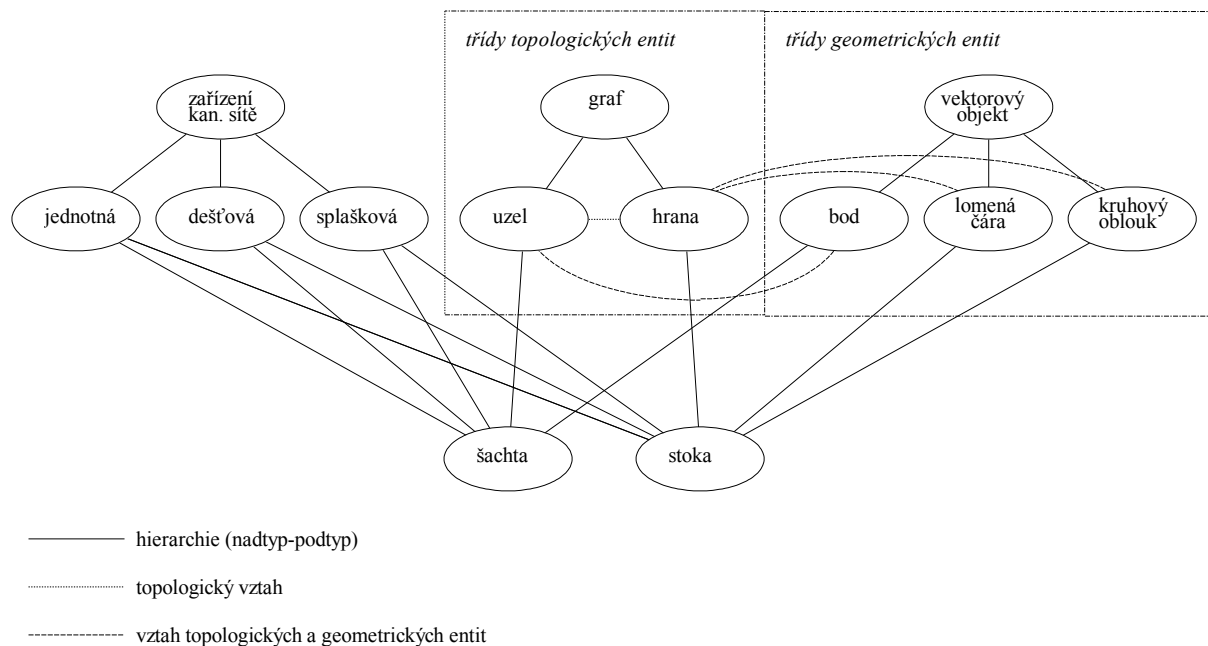
² Binární relace $r \subseteq M \times M$ je acyklická, pokud platí: je-li $n \geq 2$, $m_1, m_2, \dots, m_n \in M$, $(m_1, m_2), \dots, (m_{n-1}, m_n) \in M \times M$, pak $m_1 \neq m_n$.

- b) třída lomená čára
- 2. třída rastrových dat
 - a) třída pixel
 - b) třída řetěz pixelů
 - c) třída oblast pixelů
- II. předmětný objekt

Příklady topologických relací:

- objekt typu hrana začíná v objektu typu uzel a končí v objektu typu uzel
- objekt typu prostorový objekt leží uvnitř objektu typu areál
- objekt typu hrana tvoří hranici mezi dvěma objekty typu areál
- objekt typu prostorový objekt leží {nalevo, napravo} od objektu typu hrana

Příklad hierarchie jednoduchého datového modelu:



Obr. 24 - Příklad struktury jednoduchého datového modelu GIS

3. Rastrový a vektorový model prostorových dat

Základními modely prostorové složky geografických dat jsou rastrové a vektorové modely. Rastrový model dělí modelovanou oblast (WOI) do pravoúhlé sítě buněk, každá buňka obsahuje jednotlivou hodnotu sledovaného údaje. Buňce odpovídá plocha sledované oblasti (území). Jedna množina buněk spolu s asociovanými hodnotami vytváří vrstvu. V databázi může být uloženo mnoho vrstev. Vektorový model využívá pro vymezení lokalizace diskrétní bodové nebo liniové elementy - úsečky. Diskrétní objekty vektorového datového modelu vznikají spojením úseček. Na rozdíl o rastrového datového modelu nemusí vektorový datový model pokrývat celé území. (Rastrový model říká co se vyskytuje kdekoliv - v každém místě oblasti, vektorový model říká, kde se cokoliv vyskytuje, zobrazuje umístění každého objektu.)

V obou modelech je prostorová informace reprezentována základními (homogenními) jednotkami. V rastrových modelech jsou to pixely (buňky), ve vektorových modelech jsou homogenními jednotkami body, lomené čáry a polygony. Pokud rastrový a vektorový model popisují shodné území ve srovnatelné přesnosti, je v rastrovém modelu řádově více základních jednotek než ve vektorovém modelu. Rozměr základních jednotek v rastrovém modelu je konstantní, rozměr základních jednotek ve vektorových modelech je velice proměnlivý - lomená čára se může skládat ze dvou bodů nebo také z několika desítek tisíc bodů.

Rastrový model

Výhody

- jednoduchá datová struktura - snadné provádění některých operací - například mezi překrývajícími se vrstvami (overlay) - viz popis analytických funkcí
- zachycuje data s proměnlivou prostorovou složkou (barevný digitální obraz)
- nemusí být žádná primární znalost jevu pro získání dat, vzorkování lze provést automaticky

Nevýhody

- málo kompaktní (řídce matice)
- obtížná reprezentace topologických vlastností
- neestetický výstup reprezentovaných dat (pouze za cenu zvýšení hustoty dat)
- nemožnost libovolného zvětšování obrazu

Vektorový model

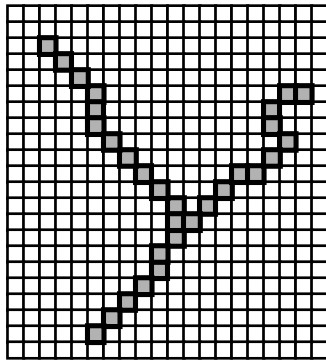
Výhody

- kompaktní datová struktura
- výhodné uchování topologických vlastností
- blíží se ručně kresleným mapám, možnost estetického výstupu dat

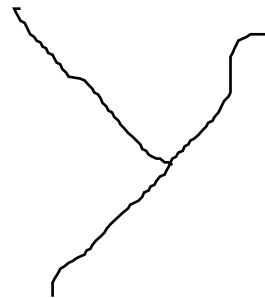
Nevýhody

- složitější datová struktura
- obtížnější overlay operace
- nevýhodná reprezentace dat s proměnlivou prostorovou složkou (digitální obraz)

rastrový model



vektorový model



Obr. 25 - Příklad rastrové a vektorové reprezentace

Volba datového modelu prostorových dat a datového modelu popisných dat je dána následujícími požadavky na vlastnosti implementace GIS:

- stupněm integrace geografického informačního systému do "vnějšího" informačního systému (datový model GIS musí být přizpůsoben okolí, pokud je GIS komponentou širšího informačního systému)
- základní architekturou informačního systému - centralizovaný x distribuovaný systém (komunikace v distribuovaném systému je složitější a tedy nevýhodná pro komplikované datové modely)
- cenou pořízení dat a cenou údržby dat ve zvoleném modelu (pořízení i údržba dat ve složitějších modelech je dražší)
- softwarovým vybavením, které je k dispozici
- zkušenostmi a znalostmi klíčových pracovníků
- historickými precedenty v dané oblasti

Volba mezi rastrovým nebo vektorovým GIS není řešením otázky, který je lepší, ale otázky, za jakých podmínek je který z nich výhodnější a jakou máme možnost využít nevhodnější řešení v konkrétních případech.

Lze diskutovat o čtyřech oblastech:

- přesnost souřadnic
- rychlost analytických funkcí

- požadavky na velikost paměti
- charakteristiky jevu, který má GIS modelovat

Přesnost souřadnic

- u rastru závisí na přesnosti pořízení dat (družicového snímku, skenovacího zařízení), obtížně se reprezentují liniové objekty s malým průměrem
- u rastru není často známo, jaký vztah má poloha hraničních bodů nebo středu buňky k reálnému kartografickému souřadnému systému
- u vektorových dat může být přesnost prakticky libovolná, závisí pouze na metodě reprezentace souřadnic (počet bitů na jednu souřadnici) a na velikosti zobrazovaného území, velká přesnost však často nemá smysl, je nutné zohlednit vlastnosti vstupujících dat
- přesnost vektorových dat se uplatní při zpracování dat z přesných geodetických měření, při vytváření map z těchto měření
- některé jevy (hranice parcel, administrativních jednotek, osy vedení inženýrských sítí) jsou vhodné pro reprezentaci vektory s velkou přesností, u některých jevů (hranice druhů lesních porostů, rozšíření živočichů a pod.) jsou přesné hranic zavádějící

Výpočetní rychlost

- rastrová data mohou být velmi rychle zpracována při odpovědích na většinu analytických dotazů včetně překryvů (vrstev), blízkosti, často stačí porovnání buněk v různých vrstvách
- ve vektorových systémech je pro stejné odpovědi vyžadována větší výpočetní kapacita - musí být řešeny komplexní geometrické problémy například výpočet polohy a průsečíku dvou úseček
- při výpočtech průniků polygonů vznikají singulární polygony

Požadavky na paměť

- v nejjednodušším režimu vyžadují rastrová data na jednu buňku jeden nebo dva byty, v těchto případech je podstatně omezen počet použitých řádků a sloupců
- rastrová data je možné komprimovat několika metodami, stupeň komprese závisí na prostorové proměnlivosti dat
- u vektorových dat jsou nároky na paměť malé, závisí na složitosti objektů a přesnosti souřadnic

Charakteristika modelovaného jevu

- rastrovou reprezentaci je možné použít tam, kde není apriori známa prostorová variabilita jevu
- typická data v rastrovém formátu pocházejí ze satelitních nebo leteckých snímků a ze skenování
- vektorová reprezentace poskytuje lepší možnosti, pokud se prostorová variabilita liší v různých oblastech území
- typická data ve vektorovém formátu pocházejí z geodetických měření, zobrazují socio-ekonomické jevy, objekty utilit, vlastnické hranice, apod.

4. Data v rastrovém modelu

Hodnoty uchované v buňce (pixelu)

- hodnota uchovaná v buňce (pixelu) závisí jednak na kódované realitě, jednak na vlastnostech SW
- možnosti: celá čísla, čísla v pohyblivé čárce, alfanumerické hodnoty
- každá buňka - pixel - má pouze jednu hodnotu, což nepřesně interpretuje realitu (hranice jevu prochází buňkou), některé systémy umožňují uchovat více hodnot v jedné buňce a jejich podíly
- rastrová data mohou být vizualizována jako množina mapových vrstev (mapová vrstva je množina dat popisující jednoduchou charakteristiku pro každou buňku uvnitř sledované oblasti)
- typické rastrové databáze obsahují více než sto vrstev, důležitou charakteristikou vrstvy je rozlišení (rozměr buňky), orientace (azimut, úhel mezi severem a směrem sloupců rastru)

5. Data ve vektorovém modelu

5.1 Špagetový (spaghetti) model prostorových dat

Pojem *špagetový model vektorových dat* je odvozen od způsobu digitalizace papírové mapy a uchování digitalizovaných dat. Papírová mapa je převedena po čarách do posloupností bodů definovaných souřadnicemi X a Y - struktur typu lomená čára. Bodové prvky jsou převedeny do dvojic

souřadnic - struktur typu bod. Soubor lomených čar a bodů nemá žádnou vnitřní strukturu - odtud termín špagetový model.

V širším pojetí lze pojem špagetový model použít při charakteristice datového modelu, ve kterém jsou objekty prostorových dat uchovány bez vazeb na jiné objekty prostorových dat stejné nebo jiných tříd (například v sekvenčním souboru typu DXF nebo DGN).

Ve špagetovém datovém modelu tedy nejsou topologické objekty. Špagetový model nepostihuje relace mezi prostorovými daty, vztahy musí být odvozeny výpočtem. Špagetový model je jednoduchá struktura vhodná pro počítačovou reprodukci map, ale i pro uchovávání prostorových dat, se kterými nebudou prováděny prostorové operace a které mají charakter pozadí - podkladu pro práci s jinými prostorovými daty. Výhodou špagetového modelu je levná údržba prostorových dat.

5.2 Topologický model prostorových dat

Topologický model je nejrozšířenějším modelem, který uchovává prostorová data a vztahy mezi nimi. Tímto modelem lze reprezentovat areály i sítě. Základními objekty v tomto modelu jsou:

- rovinný graf (arc - node model),
- areálový rovinný graf
- síťový graf s identifikovanými hranami
- síťový graf s identifikovanými uzly

Topologický model obsahuje topologické objekty, nemusí však obsahovat geometrické objekty (v tomto málo častém případě probíhá údržba prostorových dat bez jejich grafického vyjádření).

Mezi hlavní výhody topologického modelu je možnost provádět analýzy prostorových dat bez užití souřadnic - bez práce s geometrickými objekty. Nevýhodou je nutnost údržby topologické struktury, která se často - zejména pro grafy s velkým počtem prvků - provádí dávkově.

6. Základní entity prostorových dat

6.1 Geometrické entity

Bod

```
typedef struct
{
    long    x; // souřadnice x
    long    y; // souřadnice y
}
pointT;
```

příklady: střed sídla, definiční bod budovy, šachta kanalizační sítě

Lomená čára

- posloupnost bodů
- efektivní reprezentace lomených čar: přírůstky (offset) ve vektorech, Freemanův řetězový kód v rastroch
- příklady: osa komunikační sítě, osa koleje, osa kanalizační sítě, vodovodní sítě, plynovodní sítě, elektrické sítě, vodního toku

Polygon, areál

- polygon - uzavřená lomená čára
- areál - posloupnost orientovaných polygonů
- příklady: hranice fyzikogeografických oblastí (typy půdy, lesa, užití Země), hranice administrativních jednotek (okresy, obce, katastrální území)

Sledované území může být pokryto plochami (areály) dvěma způsoby:

1. entity jsou izolované plochy, mohou se překrývat (příklad: plochy zasažené různými lesními požáry), prostor nemusí být plně pokryt entitami
 2. každé místo patří právě jedné entitě (až na body hraničních čar), plochy plně pokrývají prostor
- První způsob pokrytí území může být konvertován do druhého.

Z toho vyplývají dvě možnosti uchování ploch:

- polygony - každý polygon je uložen jako posloupnost bodů, většina bodů je uložena dvakrát
- diagraf rovinného grafu (arc-node model) - jsou uloženy hranice mezi areály, každý jen jednou s ukazateli na areály

Pixel

- primitiv – základní jednotka rastrových dat, důležitá je jeho velikost

Řetěz pixelů

- řetěz pixelů (např. ve Freemanově kódu) vymezuje liniový prvek v rastrovém modelu
- příklady: osa liniového prvku (inženýrské sítě, komunikace), skelet liniového prvku

Oblast pixelů

- množina pixelů – vymezuje areálový prvek v rastrovém modelu
- příklady: fyzickogeografické oblasti, oblasti land-use

6.2 Topologické entity

Definice rovinného grafu

Orientovaný graf (U, H) je binární relace H na konečné množině U (H je tedy množina dvojic prvků z množiny U). Prvky množiny U nazýváme uzly (uzlové body), prvky množiny H hrany grafu. Grafy je běžné zaznamenávat pomocí kresby, pro tuto kresbu budeme používat termín *diagraf*. *diagrafem* grafu (U, H) budeme rozumět množinu bodů euklidovské roviny, které vzájemně jednoznačně korespondují množině U a množinu linií (oblouků) vzájemně jednoznačně korespondující množině H . Platí přitom, že žádná z linií jako svůj vnitřní bod neobsahuje bod korespondující s některým z uzlů a že linie odpovídající hraně tvořené dvěma uzly má za koncové body ty body, které s uzly korespondují. *Stupeň uzlového bodu* je definován jako počet hran, se kterými uzlový bod inciduje (počet dvojic - prvků množiny H , ve kterých se uzlový bod vyskytuje).

Graf je rovinný, pokud existuje jeho diagraf takový, že linie diagrafu korespondující hranám grafu se protínají právě jen v uzlových bodech.

Rovinný graf nazveme *rovinným areálovým grafem*, pokud jeho hrany tvoří hranice areálů v rovině. Každá hrana grafu je identifikována dvěma údaji: jmény (identifikacemi) areálů, jejichž hranici tvoří.

Datové struktury areálového grafu

Hrany

```
typedef struct
{
    long        numVert; // počet vrcholů
    pointT      *vert;   // vrcholy
    graphNodT   *ub1;    // počáteční uzlový bod
    graphNodT   *ub2;    // koncový uzlový bod
    idT         id1;     // identifikace levého areálu
    idT         id2;     // identifikace pravého areálu
}
graphEdgeT;
```

Uzly

```
typedef struct
{
    pointT      node;    // souřadnice uzlového bodu
    int         degree;  // stupeň uzlového bodu
    graphEdgeT  edge[]; // hrany incidující s uzlovým bodem
}
```

```

    }
    graphNodT;

```

Hranice areálu

```

typedef
{
    long          numEdge; // počet hran tvořících hranici
    graphEdgeT   edge[]; // seznam hran tvořících hranici
}
boundaryT;

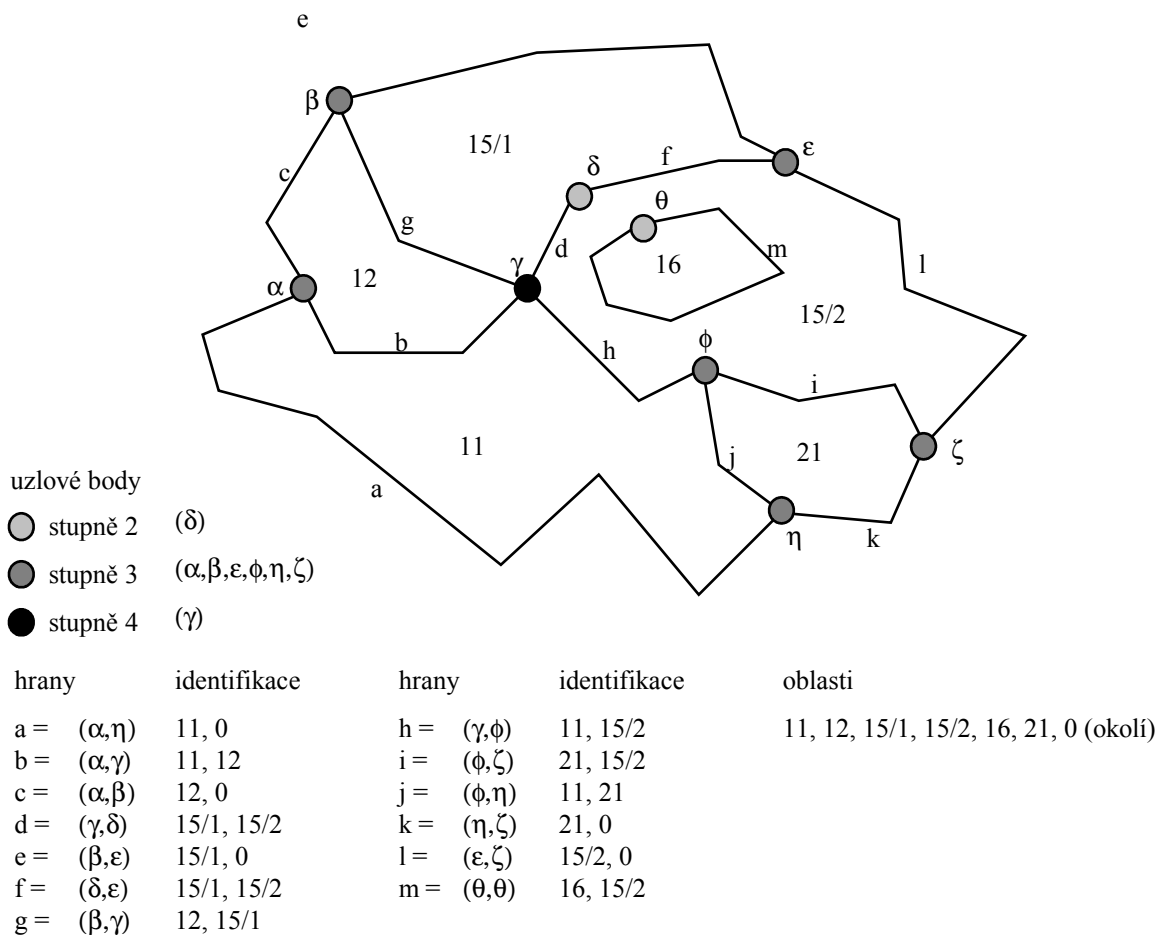
```

Areál

```

typedef struct
{
    idT          id;          // identifikace areálu
    long         numBoundary; // počet hranic tvořících areál
    boundaryT    boundary[]; // seznam hranic tvořících areál
}
areaT;

```



Obr. 26 - Příklad diagrafu rovinného areálového grafu

Příklad - odvození rovinného grafu ze špagetového modelu

Strukturu rovinného grafu lze vytvořit výpočtem ze špagetového modelu podle umístění prvků tvořících graf v prostoru (z jejich geometrie). V tomto případě je nutné provést následující kontroly vlastností rovinného grafu:

- kontrola stupňů uzlových bodů (pokud je graf areálový, nesmí se v grafu vyskytovat uzlové body stupně 1)
- kontrola křížení hran (hrany se musí křížit jen v uzlových bodech, v těchto bodech musí být hrany ukončeny)
- kontrola identifikací hran a uzlů (pokud jsou prvky tvořící graf jednoznačně identifikovány, je nutné kontrolovat správnost těchto identifikací)
- vyznačení potenciálních chyb v uzlových bodech a jejich oprava (v některých případech je vhodné zjistit, zda dva různé uzlové body nejsou příliš blízko, nebo zda uzlový bod není příliš blízko hraně grafu, se kterou neinciduje - testovaná vzdálenost je parametrem kontroly)
- vyznačení a oprava singularit v grafu (hrany složené ze dvou nebo tří bodů, které mají shodný počáteční a koncový bod)

7. Vztahy mezi prostorovými daty

Tři typy relací:

1. relace, které jsou použity pro konstrukci komplexních objektů z primitivů: posloupnost bodů tvoří lomenou čáru, posloupnost polygonů tvoří plochu (areál)
2. relace, které lze spočítat z geometrie (ze souřadnic objektů): průsečík dvou lomených čar je uzlem grafu, bod leží v areálu
3. relace, které nelze spočítat ze souřadnic, musí být pořízeny při vstupu dat: ze 2D geometrie nelze spočítat uzly kanalizační sítě

Příklady prostorových relací

- bod-bod: vzdálenost mezi body, nejbližší bod k zadanému bodu
- bod-čára: bod je vrcholem lomené čáry, vzdálenost bodu od čáry, nejbližší bod k čáře
- bod-areál: bod leží v areálu, areál může být vidět ze zadaného bodu
- čára-čára: křížení čar, čára leží na čáře, "vtéká do" koncový bod jedné čáry leží na jiné
- čára- areál: čára kříží areál, čára je hranicí areálu
- areál - areál: množinové operace s areály (průnik, rozdíl, sjednocení), vzdálenost areálů, sousednost areálů

8. Zkušební otázky

1. Porovnej vlastnosti, výhody a nevýhody vektorových a rastrových dat
2. Jaké znáš reprezentace plošných objektů v datových modelech prostorových dat?
3. Souřadnice X a Y v kartografickém souřadném systému JTSK mají deset cifer. Navrhni úspornější reprezentaci lomené čáry v tomto systému, než je posloupnost bodů s těmito souřadnicemi. Proveď diskusi výhod a nevýhod svého návrhu.
4. Uveď příklad geografických dat, pro která je vhodná vektorová forma uchování. Uveď příklad geografických dat, pro která je vhodná rastrová forma uchování.
5. Uveď další příklady prostorových relací k těm, které jsou uvedeny v kapitole 7.

Připomínky a dotazy k obsahu lekce posílej, prosím, na adresu:
Rudolf Richter, richter@fi.muni.cz